Author:
Gallo Cordoba, Beatriz Eugenia

Title:
Statistical Methods for Investigating the Ethnic Achievement Gap in Colombia

Statistical Methods for Investigating the Ethnic Achievement Gap in Colombia

Beatriz Eugenia Gallo Cordoba

School of Education
University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Social Sciences and Law.

November 2019

Word count: 79,885

Abstract

In industrialised countries like the US and the UK, the substantial body of literature about the problem of ethnic achievement gaps has informed a series of policies attempting to tackle it. Multiple methods have been applied, often borrowed from disciplines that have refined them to predict academic achievement or examine school effectiveness. In Colombia, only until 2018, the ethnic achievement gap became part of the policy agenda, but there is only one study describing the gaps, offering limited evidence to inform policy. This thesis views the ethnic achievement gap as a multilevel phenomenon of averages (reflecting the effect of student, school and local authority (LA) characteristics) and has both substantive and methodological aims. The substantive aim is to improve the current understanding of ethnic achievement gaps in Colombia (between White-mixed European descendants, the largest ethnic group- and Afrocolombian, Indigenous, and other ethnic minorities), including how they vary over time, within different subpopulations, and across different levels of the education system: students, schools and LAs. The methodological aim is to study the potential of tailored measurement and statistical modelling techniques for improving the understanding of the nature of ethnic achievement gaps. The analysis applies single-level and multilevel modelling, as well as multivariate analysis methods to the results of a nation-wide compulsory examination for 11th grade (age 16/17) students between 2008 and 2013. Substantively, it is shown that the ethnic achievement gap of each minority group reflects different underlying processes at each level of the education system. Methodologically, it is argued that multilevel modelling is preferable to single-level modelling in some, but not all circumstances and each measurement method impacts the analysis of ethnic achievement gaps in different ways. The implications for policy and the consolidation of methods that contribute to the understanding of ethnic achievement gaps are discussed.

To my beloved family

A mis ñocos que amo todo mi ser

Acknowledgements

I feel incredibly fortunate to have found the support of many people throughout the path to completing this dissertation.

My supervisors, George Leckie and Bill Browne, were the best team I could have asked for. I sincerely appreciate your guidance and support, and the time you dedicated to commenting on my work.

David Manley and Steve Strand, members of my examination panel, made my *Viva Voce* experience incredibly constructive and enjoyable. I appreciate the time for examining my thesis and the advice and different insights that we discussed during our meeting.

Sally Thomas and Harvey Goldstein, members of my progression panel, encouraged me to carefully reflect on the contribution of my thesis and its pertinence for policy.

Jo, Ann, Karen and Kevin, from the EU Exit & Strategy Team at the Welsh Government made my apprenticeship with them an enriching, rewarding and exciting experience that helped me shape the way I communicate the messages in this thesis.

My colleagues at room 3.03 provided treats from around the world, support and inspiration. I would like to especially thank Tanya, Diana, Dulce, Trang, Claudio, Maria Lucía, Aliandra and Margarita for reading different versions of this document and for kindly offering your guidance and expertise when I needed it. Besides, Caro, Merve, Trang, Saba, Satpal and Shimmi read different chapters of my thesis and challenged me with their questions during my mock viva.

My flatmates have tolerated my odd working hours and have been an invaluable source of encouragement. Chiara, Fanny and Lina, thank you for our discussions at the kitchen and unlimited emotional support.

My software choice has Cienfi's trademark. I would like to thank Julio César Alonso for introducing me to R, LaTeX and Mendeley, and for all the learning opportunities that equipped me for my PhD. Similarly, Blanca Zuluaga and Juan Esteban Carranza introduced me to different perspectives and ideas about academic research.

My parents skipped 13 generations in social mobility to allow me to be here, and despite the over 8,000 km of distance, they and my sister always made me feel close to home. It is a privilege to have you as my family. Maria and Nico became my source of pampering in Europe and made it easier to be away from home in the toughest moments.

This work was supported by the Economic and Social Research Council [grant number ES/J50015X/1] through the Advanced Quantitative Methods Pathway at the South West Doctoral Training Partnership.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

A related paper whose findings are not reported in this thesis was published as: Gallo Córdoba, B. (2017). Using social equity theory to explain ethnic achievement gaps in developing countries: Evidence from Colombia. *TOR*, (3), 12-18. Retrieved from `http://www.swdtp.ac.uk/files/2017/05/TOR-Issue-3-May-17-Final.pdf`.

SIGNED: .... DATE:..........................

Contents

List of Tables

List of Figures

Acronyms

AIC Akaike information criterion.

BC Buscando Colegio.

BIC Bayesian information criterion.

Buscando Colegio It literally translates "searching for school". This is the database that contains venue-level information about all colombian schools.

Colombian National Assessment Institute The national-level institution that is in charge of educational assessment in Colombia, including the administration of SABER 11 exam.

DANE National Bureau of Statistics.

DNP Colombian Department for National Planning.

ERIC Education Resources Information Center.

ESRC Economic and Social Research Council.

FA factor analysis.

FSM Free School Meal.

ICC intraclass correlation.

Icfes Colombian National Assessment Institute.

IEA International Association for the Evaluation of Educational Achievement.

LA local authority.

LCA latent class analysis.

LTA latent trait analysis.

Ministry of Education Colombian national-level organisation that is responsible for education.

MoE Ministry of Education.

NAEP National Assessment of Educational Progress.

National Development Plan A four-year plan outlining the central government programs and investments in order to reach development goals.

NCES National Centre for Education Statistics in the USA.

NDP National Development Plan.

OECD Organisation for Economic Co-operation and Development.

OLS Ordinary Least Squares.

PCA principal component analysis.

PIRLS Progress in International Reading Literacy Study.

PISA Programme for International Student Assessment.

SABER 11 Exam that all Colombian students must present at the end of compulsory education.

SD standard deviations.

SES socio-economic status.

Sisben System for the Selection of Social Programs' Beneficiaries.

System for the Selection of Social Programs' Beneficiaries An instrument for selecting public policy beneficiaries in Colombia.

TIMSS Trends in International Mathematics and Science Study.

VCM variance component model.

1 | Introduction

1.1 Research Problem

Around the world and especially in industrialised countries such as the United States (US) and the United Kingdom (UK), education, economics, sociology, human geography and other social science disciplines contribute to research on ethnic disparities in academic achievement. As reviewed in chapter 2, in these contexts, the debate about the nature of the gaps and the extent to which they can be attributed to students, schools or the broader external context has motivated, informed, influenced and criticised educational policy. Examples in the US include Coleman et al. (1966)'s report and its influence on desegregation policy and research (Card & Rothstein, 2007; Hanushek, Kain, & Rivkin, 2009; Page, Murnane, & Willett, 2008; Palardy, Rumberger, & Butler, 2015) and the more recent debate on the *No Child Left Behind Act* and the *Every Student Succeeds Act* (Fryer & Levitt, 2006, 2013; Hanushek, 1997; Lee, 2002; Quinn & Le, 2018; Rothstein, 2004). In the UK's evidence-based-policy environment, a series of government reports, including Newsom (1963)'s, Plowden (1967)'s, Taylor (1981)'s, Swann (1985)'s, Stokes, Rolfe, Hudson-Sharp, Stevens, and Strand (2015)'s and Casey (2016)'s, have paved the way for the *Race Relations Act* and its subsequent versions such as the *Equality Act 2010*. These reports and policies have also triggered academic debate and further policy changes to redistribution policy and the school accountability system in England (e.g. Demack, Drew, & Grimsley, 2000; Dustmann, Machin, & Schönberg, 2010; Leckie & Goldstein, 2019; Sammons, 2007; Strand, 2016). In contrast, in non-industrialised countries, there is little evidence about ethnic achievement gaps, and only recently, these have become a concern of public policy.

Therefore, despite the extensive research on ethnic achievement gaps, its multidisciplinary nature and focus on industrialised countries imply two gaps in the literature. First, different disciplines have favoured different methods and when measuring and modelling ethnic achievement gaps, the methods that are employed have been tailored for different applications, such as predicting academic achievement (e.g. Fryer & Levitt, 2004; Hanushek & Rivkin, 2006; Quinn, 2015a) or identifying effective schools (e.g. Plewis, 2011; Rasbash, Leckie, Pillinger, & Jenkins, 2010; S. Thomas, Sammons, Mortimore, & Smees, 1997). This implies that there is a lack of consensus about a methodological framework for the analysis of ethnic achievement gaps. Thus, tailoring the current variety of methodological approaches to study ethnic achievement gaps can provide further insights about them. Second, the current understanding of ethnic achievement gaps is built mostly on evidence for industrialised countries, which does not necessarily apply to other settings. Therefore, providing evidence from a less explored context contributes to improving the evidence base for the construction of a more general understanding of ethnic achievement gaps around the world.

One of these rarely explored countries is Colombia. As described in chapter 3, Colombia is a middle-income country with three main ethnic groups: White (mixed European descendants, the largest ethnic group), Afrocolombian (descendants of slaves brought during colonial times) and Indigenous (pre-colonial inhabitants). These two minority groups have been described as socioeconomically disadvantaged, with lower school enrolment rates, average salaries, and the likelihood of securing formal employment, in comparison to the White majority (del Cairo & Jaramillo Marín, 2008; Naciones Unidas, 2009; Pérez Marulanda & Mora, 2014; M. Rangel, 2001; Romero Mora, Genes Díaz, & Ospino Meriño, 2017). Nevertheless, there is only one study exploring ethnic achievement gaps in the country. Sánchez-Jabba (2011) estimated that in 2010, 11th grade (age 16/17) minority (Indigenous and Afrocolombian, pooled together) students scored around 0.3 standard deviations (SD) lower than White students. This gap is smaller than the 1.2 SD and 0.89 SD Black-White gap for US 8th grade (age 13/14) students in 1971 and 2011, respectively (Page et al., 2008), and than the 0.45 SD gap between White British and Black Caribbean Year 6 (age 11) students in England in 2008 (Strand, 2014b). Nonetheless, a gap of this magnitude is still substantial and needs addressing from a policy perspective, as the Colombian government recognised for the first time in 2018a. However, Sánchez-Jabba (2011)'s evidence on its own is insufficient to formulate a well-informed public policy. For example, it only provides evidence for one cohort of students and, crucially, it assumes that the patterns of underachievement are the same for Afrocolombian and Indigenous students.

Therefore, studying ethnic achievement gaps in Colombia contributes to the international literature on ethnic achievement gaps, as well as to policymaking in the country. Internationally, the Colombian context is an interesting case study since lessons can perhaps be learned from a less unequal context. Furthermore, the methodological challenges that arise by examining a different setting provide an opportunity to revisit existing methodological approaches. In Colombia, improving the evidence base of ethnic achievement gaps can improve the pertinence of the policies that aim to reduce ethnic inequality.

With these considerations, this thesis has both substantive and methodological aims. Substantively, the aim is to improve the current understanding of ethnic achievement gaps in Colombia, including how they vary over time, within different subgroups (e.g., boys and girls, socioeconomically advantaged and disadvantaged, private and state schools), and across different levels of the education system (students, schools and local authorities (LAs)). The methodological aim is to study the potential of problem-specific measurement and statistical modelling techniques for improving the understanding of the nature of ethnic achievement gaps. Both aims are addressed using nation-wide administrative data of students taking a compulsory exam (called SABER 11) for cohorts of 11th grade (age 16/17) students between 2008 and 2013, which are described in chapter 4. These data are analysed using single- and multilevel modelling and techniques derived from mediation analysis, in a pragmatic approach described in chapter 5.

This introductory chapter presents the roadmap of the thesis. The following section presents the overarching research questions that are addressed in the four research chapters of the thesis (chapters 6, 7, 8 and 9) and how addressing them contributes to the ethnic achievement gaps methodological and substantive literature and policy. Section 1.3 presents the methodological contribution and rationale for the thesis as a whole. Section 1.4 delineates the scope of this thesis. Finally, section 1.5 presents the outline of the thesis. Each research chapter provides a detailed discussion of the gap in the literature it addresses, its motivation and contribution.

1.2 Research Questions

In this thesis, the ethnic achievement gap is understood as a multilevel "phenomenon of averages" (Rothstein, 2004, p.5), which recognises that differences in average test scores between White and minority students (overall ethnic achievement gaps) mirror information about different levels of the school system (student, schools and LAs), their characteristics and possible subgroups. The research questions in this thesis interrogate the methods to uncover this information and the substantive findings for the Colombian context.

Specifically, this thesis aims to answer four overarching research questions:

1. **How to decompose the overall ethnic achievement gaps into differences within and between schools and LAs? What are the results of applying this technique to the Colombian context? - Chapter 6**

   Ethnic achievement gaps have been attributed to differences in student, school and higher-level (e.g. LA) characteristics, as explored in the literature review (chapter 2), leading to different policy recommendations and accountability claims (Herman, 2009; Jones, Toma, & Zimmer, 2008; Mohammadpour & Ghafar, 2014; Voight, Hanson, O'Malley, & Adekanye, 2015). One way of understanding how these levels contribute to the achievement gaps is to decompose them into their within-school, between-school/within LA and between LA components. Nonetheless, the method chosen for this decomposition is controversial, and its pertinence for policy is limited. Chapter 6 extends this decomposition methodology to improve its relevance for policy. Its application to the Colombian context illustrates the additional insights about the ethnic achievement gaps and potential policy options that are uncovered thanks to this extension.

2. **How do the different index construction techniques for operationalising socio-economic status (SES) used in the literature affect the estimates of the conditional ethnic achievement gaps and their components at each level of the education system? - Chapter 7**

   In the ethnic achievement gap literature, SES is considered an important explanation for the existence of such gaps (e.g. Blanco, 2017; Coleman et al., 1966; Strand, 2016). In practice, data constraints often determine the operationalisation of SES, without consensus about the best possible approach to this problem (Buchmann, 2002; Sirin, 2005; White, 1982). With the growing data availability and the desire to estimate more realistically complex models, the creation of composite indicators is a sensible option for operationalising SES (Johnston, Jones, & Manley, 2018). There are many different index-construction techniques. Nonetheless, the existing advice for researchers about this measurement decision does not consider these alternatives or how choosing amongst them may affect the estimation results and the conclusions drawn by the studies. Chapter 7 examines the effects of using alternative index-creation methods to operationalise SES to study ethnic achievement gaps and proposes a framework for analysing the effects of SES operationalisation.

3. **To what extent do student, school and LA characteristics explain the overall ethnic achievement gaps and their components at each level of the education system? How do single-level and multilevel modelling results compare when answering this question? - Chapter 8**

   The overall ethnic achievement gaps are average differences that result from the combination of multiple social processes operating at different levels in the school system (student, schools and LAs) and society. Identifying these variables and levels is critical to uncover potential policy paths to reduce ethnic achievement gaps. The only previous study about ethnic achievement gaps in Colombia (Sánchez-Jabba, 2011), as the international literature, considered the ethnic achievement gap a student-level and not a multilevel phenomenon (e.g. Arteaga & Glewwe, 2019; Bali & Alvarez, 2004; Blanco, 2017; Condron, 2009; Cook & Evans, 2000; Grogan-Kaylor & Woolley, 2010; Santibañez, 2016; Stiefel, Schwartz, & Gould Ellen, 2007; Wilson, Burgess, & Briggs, 2011). Consequently, it is unknown how these different levels of the school system interrelate with student, school and LA characteristics to produce ethnic achievement gaps, and what is the most appropriate modelling approach to analyse this problem. Chapter 8 evaluates the advantages and disadvantages of using single-level and multilevel models to approach this problem and accordingly applies the methods to study the ethnic achievement gaps in Colombia.

4. **Are ethnic achievement gaps uniform across subgroups of students, schools and LAs? – Chapter 9**

Exploring how the ethnic achievement gaps vary across subgroups of students (e.g. boys and girls, students with high and low SES), schools (e.g. state and private) and LAs (e.g. with high and low fiscal performance) allows for more granularity in the understanding and policy recommendations about ethnic achievement gaps. Identifying the level of the education system (students, schools or LAs) at which these variations happen refines the insights about the processes behind this variation. In the international literature, the few studies that have examined variation in the gap among subgroups of students and schools (e.g. Chatterji, 2006; Strand, 2014b, 2016; S. Thomas et al., 1997) normally focus on the overall gap and not on its components at different levels of the school system. Chapter 9 incorporates the understanding of the ethnic achievement gaps as a multilevel phenomenon to the analysis of fixed (interaction terms) and random (random slopes) variation in the ethnic achievement gaps. It also examines whether those subgroups with narrower gaps are also the subgroups with higher expected achievement.

Each research question is addressed in a different chapter (chapters 6 to 9), which discusses in more detail the gap in the literature that it addresses. The following section briefly summarises the importance of these research questions and how they present original findings.

1.3 Contribution and Rationale

Research questions 1 and 2 are mainly methodological, while research questions 3 and 4 are mainly substantive. Nonetheless, there are both methodological and substantive considerations in each of the research chapters (6 to 9) that elaborate on these questions. As a whole, addressing this set of questions contributes to the methodological and substantive ethnic achievement gap literature and to inform public policy in Colombia.

Methodologically, this thesis develops a framework for analysis that has been tailored to the study of ethnic achievement gaps, understood as a multilevel phenomenon. This framework consists of a series of steps to guide the analysis of ethnic achievement gaps while understanding the effect of methodological choices on the research results. These steps can guide the analysis of ethnic gaps in different contexts (beyond Colombia) and for different outcomes. By doing so, this thesis benefits applied researchers and the broader academic community with recommendations to increase transparency in research results, by recognising the extent to which subjective decisions impact research findings. Simultaneously, this framework contributes to a more detailed understanding of the ethnic achievement gap.

Substantively, this thesis contributes to a more general understanding of ethnic achievement gaps by examining the ethnic achievement gaps as a multilevel phenomenon, providing further insights into student, school and LA processes behind the gap. Additionally, this thesis contributes to the substantive achievement gap literature by examining Colombia, a country with only one previous study about ethnic achievement gaps (Sánchez-Jabba, 2011). The particularities of the Colombian context, described in chapter 3 provide a different perspective in comparison to the industrialised settings that the literature has scrutinised so far.

Finally, this thesis provides timely evidence to inform the new National Development Plan in Colombia (Congress of Colombia, 2019), that for the first time recognises the need to address ethnic achievement gaps, among other inequalities among ethnic groups.

1.4 Scope of the Research

This thesis presents an exploratory correlational analysis of ethnic achievement gaps at the end of high-school in Colombia between 2008 and 2013. The literature on achievement gaps is not limited to ethnicity but also considers gender, age, SES, prior achievement and other characteristics (Crawford, 2007; Georges & Pallas, 2010; Ma, Yuan, & Luo, 2016; Strand, 2014a; S. Thomas et al., 1997). Unlike ethnic achievement gaps, gender and SES achievement gaps have received much more attention in the Colombian literature (e.g. de Jorge-Moreno, Díaz Castro, Rodríguez Vega, & Segura Gutiérrez, 2018; Gamboa, Casas, & Piñeros, 2003; Muñoz Wilches, 2010; Orjuela, 2013). Concerns about the relationship between achievement and these student and school characteristic are secondary in this thesis since the focus is on ethnic achievement gaps in Colombia. When these additional student and school characteristics are considered, the interest is in their relationship with ethnic achievement gaps, and not on the causal mechanisms behind these relationships or their ability to predict academic achievement or measure school quality, for example.

Throughout this thesis, the studied measure of academic achievement is limited to maths test scores. This decision does not intend to neglect the importance of achievement in other subject areas, such as language or science, and in other educational outcomes, such as citizenship, critical thinking, communication skills, social relationships or personal well-being (Oreopoulos & Salvanes, 2011). Using maths test scores improves the comparability of the results with those of other education systems for two reasons. First, maths is usually a compulsory subject even in education systems with tiering and other selection practices (such as the UK and the US), which means that the results of this thesis can be compared with those of students of the same educational stage as Colombian students taking the SABER 11 exam. Second, the contents of maths tests are more likely to be comparable across (western) education systems (Kyriakides, 2004). Since this thesis aims to contribute to a more general understanding of the ethnic achievement gaps, producing results which can be compared to those in other countries is crucial.

Finally, this thesis does not claim to question every possible methodological decision in the study of ethnic achievement gaps. For example, alternative model-building algorithms (e.g. bottom-up or top-down) or estimation methods (e.g. Markov Chain Monte Carlo or Iterative Generalized Least Squares) are not considered. Consequently, the framework for analysis that is proposed is by no means an exhaustive list of all the decisions researchers make but should be viewed as a general framework that can (and should) be tailored to each research context.

1.5 Outline of the Thesis

This thesis can be broadly divided into three parts. The introductory chapters, from chapter 1 to chapter 5, contextualise the thesis by providing an overview of the literature, the Colombian context, the data and methods that are used through the thesis. The contribution to knowledge of this thesis is presented in the research chapters (chapters 6 to 9). Each research chapter is largely self-contained and presents the results for one of the research questions and their connections with the literature and context. Finally, chapter 10 presents an overarching discussion and conclusion of the thesis as a whole.

2 | Literature Review

2.1 Introduction

This chapter provides an overview of the explanations that multiple disciplines in the social sciences have proposed for the ethnic achievement gap in maths[1] at the end of compulsory education at each level of the education system (student, schools and higher levels) around the world. Each of the research chapters (6 to 9) contains a more specific review of the literature, according to the research question they address. The literature review in chapter 6 focuses on the methodological debate about the ethnic achievement gap decomposition. The analogous section in chapter 7 discusses the operationalisation of socio-economic status (SES). In chapter 8, it discusses the methods that have been used to quantify the importance of student, school and higher-level characteristics that explain the ethnic achievement gap. Finally, in chapter 9 it reviews the role of interactions and differential school effectiveness. The context chapter (3) also includes a specific review for educational research in Colombia, which is presented after discussing the relevant details of the Colombian education system. The methods chapter (5) also presents an overview of the relevant methodological literature.

Before delving into explanations for the ethnic achievement gap, the next section defines the main concepts in the discussion (ethnicity, achievement and SES). Then section 2.3 presents the studies that precede the current debate about ethnic achievement gaps around the world since most of the following literature attempted to challenge, complement or replicate these early developments. Section 2.4 focuses on student, school and higher-level explanations for the ethnic achievement gap. The chapter finishes with a

[1] Although not discussed in depth here, results for other subjects tend to confirm the overall patterns of the gaps in maths, but these are usually larger for language tests (Arteaga & Glewwe, 2019; Dustmann et al., 2010; Quinn & Le, 2018; Strand, 2016).

discussion of the importance of each level and a summary of this review.

2.2 Key Concepts

2.2.1 Race and Ethnicity

Race and ethnicity are used interchangeably in the achievement gaps literature (e.g. Dee, 2005; Elmelech & Lu, 2004; Sonnenschein & Galindo, 2014). However, in areas such as sociology, the use of these terms is widely debated. This thesis does not intend to add to this debate, but this section clarifies what is meant in this work when using these terms.

As Fenton (2010) discusses, the term race is usually employed to refer to a group of people that share descent, location and physical or visible characteristics. In educational research, this term is often used to refer to students from a relatively recent immigration wave (e.g. Willke, 1975). In turn, ethnicity refers to a subset of people from a nation with a shared descent and culture, and that is different from the majority group (Fenton, 2010).

In the SABER 11 database, described in chapter 4, the observed variable is the student's self-perception as pertaining or not to a minority group. These minority groups refer to groupings of Colombians with shared beliefs, practices and language. Thus, the observed variable is more closely related to the concept of ethnicity than to the concept of race, since there is no reference to the physical appearance of individuals, but to their culture. This operationalisation of the variable limits the definition of ethnicity that is adopted in this thesis and therefore, the term ethnic minority refers to those who self-identify as pertaining to an ethnic minority, as opposed to the largest ethnic group[2]. Ethnicity can be operationalised coarsely, pooling all minority groups into a single group (which is the approach followed in the only study about ethnic achievement gaps in Colombia (Sánchez-Jabba, 2011)), or more finely, differentiating among ethnic groups, which allows examining differences between minority groups, which is the case of this thesis.

2.2.2 Attainment and Achievement

According to the Education Resources Information Center (ERIC)'s thesaurus, achievement is the "level of attainment or proficiency in relation to a standard measure of per-

[2]For more information on the different ethnic groups in Colombia, please refer to chapter 3.

formance, or, of success in bringing about a desired end" (ERIC, 1966) and educational attainment means the "years of successfully completed schooling or the equivalent according to some accreditation standard" (ERIC, 1980). Nonetheless, in the UK research, the term attainment is often used to describe student performance (e.g. Connolly, 2006; Demack et al., 2000; Leckie & Goldstein, 2019; Plewis, 2011; Wilson et al., 2011), suggesting that achievement refers to success in a range of outcomes that are not restricted to test scores (Mannion, Sowerby, & I 'anson, 2015). In turn, in the US the word achievement is used almost as an equivalent of test scores (Grogan-Kaylor & Woolley, 2010; Kao & Thompson, 2003; Robinson, 2008; Wiggan, 2007).

This thesis follows ERIC's the definitions in, and reserves 'attainment' to indicate someone's maximum educational qualification obtained, as this is how parents' educational background is observed. 'Achievement' is used to refer to the degree to which students are proficient in a particular subject, measured by maths test scores.

2.2.3 Socioeconomic Status

Socio-economic status is generally understood as the position of individuals in the society, according to their access to material (e.g. food, assets) and non-material (e.g. access to a network, prestige) resources, mainly composed of factors described by parents' occupation, education, and income (Bornstein, Hahn, Suwalsky, & Haynes, 2003; Braveman, Cubbin, Marchi, Egerter, & Chavez, 2001; Finch & Hoehn, 1951; Jeynes, 2002; Rutkowski & Rutkowski, 2013; Sirin, 2005; White, 1982; Willms, 1992). This definition of SES is encompassed in this research.

As alternatives to SES, researchers often choose between two options. First, access to Bourdieu (1986)'s economic, social and cultural capital. If this is the case, measures such as the number of books or paintings in a household are usually included to operationalise cultural capital (Pokropek, Borgonovi, & Jakubowski, 2015; Yang, 2003; Yang & Gustafsson, 2004; Yang Hansen & Munck, 2012). Second, the Marxist idea of social class, which is often operationalised as parental occupation (Hart Reyes & Stanic, 1988; Muller, Riegle-Crumb, Schiller, Wilkinson, & Frank, 2010; Palardy, 2015). As further explained in chapter 7, in empirical work, data availability often determines the operationalisation and conceptualisation of SES.

2.3 Precursors of the Modern Debate

Ethnic achievement gaps are a regular and salient feature of multicultural education systems around the world. P. A. J. Stevens and Dworkin (2014) argued that research on this topic is country-specific, and its development is highly associated with each national context. Nonetheless, research on this topic has been dominated by the literature in the United States (US) and other industrialised countries.

The Coleman report is probably the most influential work in this literature. Coleman et al. (1966) showed that 12th grade (age 17/18) US ethnic minority students scored between 1 standard deviations (SD) (for Black students) and 0.05 SD (for Asian Americans) lower in maths than their White peers, with a similar pattern emerging for different subjects and school grades. Coleman et al. (1966) also showed that, after considering the student's SES, family background explained most of the variation in test scores, as 20% of the variation in minority students' test scores was associated with school differences, but this figure was only 10% for White students. Coleman et al. (1966, p.22) argued that this difference between White and minority students implied that "the average White student's achievement seems to be less affected by the strength and weaknesses of his school's facilities, curriculums and teachers than is the average minority pupil's".

In the UK, an evidence-based policy environment led to the production of a series of reports from the 1960s (Cane, 1967). The Newsom report was the first of these reports to point out the increasing immigration to London and the Midlands from the West Indies, India and Pakistan, and suggested that decisions on school segregation had to consider 'other' factors beyond education (Newsom, 1963, p.71), without explaining which these 'other' factors were. The Plowden report described the disadvantaged conditions of immigrant children and the difficulties caused by the lack of fluency in English and cultural differences (Plowden, 1967, ch. 6). They reported that parental attitudes explained 38% of the differences in achievement and were more important than differences between schools. However, this report did not estimate ethnic achievement gaps. Later, Taylor (1981) found that West Indian pupils obtained lower grades in secondary school and had a lower probability of attending tertiary education. Swann (1985) attributed these results to discrimination both, against the children at school and against their parents, which prevented them from attaining a higher SES that translated into higher academic achievement.

Heyneman (1976) and Heyneman and Loxley (1983) were pioneers in replicating the Coleman Report in non-industrialised settings, finding that schools play a relatively more important role in student achievement there than in industrialised countries. In Colombia, it was only until 2011 that Sánchez-Jabba presented evidence of a maths achievement gap of 0.35 SD between White and minority students, as further discussed in section 3.4.2.

In all these early influential studies, there is an understanding that different levels of the education system (students and schools) contribute to the existence of ethnic achievement gaps and concern for identifying the separate effects of student and school characteristics on academic achievement. The literature that followed echoed this concern, reaching for explanations at each of these levels according to the particularities of the research context.

2.4 Student- and Family-Level Explanations for Ethnic Achievement Gaps

Explanations for the ethnic achievement gap based on student characteristics can be grouped in what Conyers (2002) called deficiency theories. These theories argue that ethnic minorities are disadvantaged in different aspects, which explains their lower achievement in comparison to the majority group. Following Conyers (2002) typology, the literature has explored biological, cultural and structural disadvantages. A distinctive feature of this literature is that observational studies often include school fixed effects into their linear regression models to isolate the effects of the students from the effects of schools.

The literature claiming a biological or genetic disadvantage of ethnic minorities precedes the discussion on academic achievement and keeps growing until today (e.g. Fuerst, Lynn, & Kirkegaard, 2019; Fulk & Harrell, 1952; Herrnstein, 1994; Plomin, 2018). Authors in this area have focused on measuring and comparing the *intelligence* of ethnic groups, claiming causal relationships from correlational studies. The broad consensus is that, although there is a genetic component of academic achievement (Morris, Davies, Dorling, Richmond, & Smith, 2018), variation in this genetic component is not linked to ethnicity (Sussman, 2016). Studies about *intelligence* often face rejection from the academic community because of the racist ideas they promote and their flawed methodology (e.g. Cofnas, 2016; Comfort, 2018; Templer & Arikawa, 2006; Wicherts, Borsboom, & Dolan, 2010).

Cultural explanations assert that minority children are confronted with a different (linked to the majority group and possibly new and opposed to their own) set of values, expectations, language and norms, which hinders their academic achievement (Ogbu, 1995a). Ogbu (1995b)'s ethnographic study in the US showed that while immigrant students are more likely to interpret cultural differences as a non-threatening learning opportunity, Black students (involuntary minorities) interpret these differences as a threat for their own identity. Therefore, cultural differences would be more detrimental for Black than for immigrant students. Studies investigating the role of cultural factors frequently report a correlation between academic achievement and cultural aspects such as aspirations and plans, self-concept, attitude to school, behavioural problems, parental involvement and practices, how much parents and students value education, the student's sense of belonging to the school, or how much students value their cultural heritage (Francis & Archer, 2005; Hampden-Thompson, Guzman, & Lippman, 2013; Hartas, 2015; Knifsend, Camacho-Thompson, Juvonen, & Graham, 2018; T. Stevens, Olivárez, & Hamman, 2006; Wolfgramm, Morf, & Hannover, 2014). However, they often do not report how important these factors are for explaining the ethnic achievement gap and, when they do, do not find that these variables meaningfully explain the ethnic achievement gap (Strand, 2011; Todd & Wolpin, 2007).

Differences in language are one exception, as research has shown that in the case of first-generation immigrants, Black students in the US and minority groups who use a different language at home, differences in language skills at least partially explain the ethnic achievement gap in contexts as different as China, Germany, the US, the UK (Haag, Roppelt, & Heppt, 2015; Yang et al., 2015). Nevertheless, minority students seem to overcome this disadvantage the more they are exposed to the school system (Dustmann et al., 2010; Kieffer, 2008).

Structural explanations for the gap are based on the finding that students living with two parents tend to have a higher academic achievement than students living with only one of their parents or in other family arrangements (Cervini, Dari, & Quiroz, 2014; Heuveline, Yang, & Timberlake, 2010; Phoenix & Husain, 2007), and the observation that ethnic minorities are more likely to be monoparental than the families of the majority group. However, the importance of family structure per se has been questioned because of its strong correlation with parental styles and SES (Payne, 2003).

Precisely, SES is often recognised as the most important single explanation for the ethnic achievement gap. In the US, SES has been shown to explain between 22% and 42.3% of the Black-White achievement gap, depending on the year and school grade (Fryer & Levitt, 2004; Page et al., 2008). In other countries, it has been shown that differences in SES explain as little as 10.4% of the gap between Indigenous and non-Indigenous students in Guatemala (McEwan & Trowbridge, 2007), and as much as 78.3% of the gap between White and Pakistani students in England (Strand, 2011), as it is also the case for Colombia, as shown in chapter 8.

The variation of this result around the world is expected, given the differences of time, context and ethnic groups. However, there is also a large variation in the way in which researchers operationalise SES (Sirin, 2005; White, 1982), as shown in detail in section 7.2.2. Given the potential importance of SES, chapter 7 focuses on how the operationalisation of SES can change the conclusions about ethnic achievement gaps. As Dworkin and Turley (2014) noted, the lack of support for student-level explanations for the ethnic achievement gap even after considering differences in SES stimulated a shift towards school-level explanations for the ethnic achievement gap, which are further discussed in the next section.

2.5 School-Level Explanations for Ethnic Achievement Gaps

Three main hypotheses link schools to the ethnic achievement gap. First, that processes and practices within schools tend to favour the achievement of the majority group at the expense of ethnic minorities. Second, the interaction between peers leads to lower academic outcomes for minority students. Finally, that minority students systematically attend schools of lower quality than the ethnic majority group. The data and analysis in this thesis do not allow to differentiate between these three hypotheses. Nonetheless, the following overview provides a framework for interpreting the findings of chapters 6, 7, 8 and 9.

2.5.1 School Processes, Policies and Practices

The ethnic achievement gap may be exacerbated or alleviated by processes, policies and practices within schools. In general, more egalitarian and inclusive policies are associated with a narrower ethnic achievement gap.

One of the concepts that encapsulate these processes and practices is the school climate. Voight et al. (2015) conceptualised it as a combination of opportunities for participation, safety and connectedness and adult-student relationships. Using data of middle schools in California, they found that White, Black and Hispanic students experienced a different school climate within the same school and that these differences between ethnic groups varied between schools; there was an ethnic gap in school climate that varied between schools. Schools with a wider gap in school climate also had a wider achievement gap, and rural and socioeconomically disadvantaged schools had lower school climate gaps. Research has also shown that ethnic minorities in the US are more often taken away from the classroom because of disciplinary action, which may affect their academic achievement (Gregory, Skiba, & Noguera, 2010; Gregory & Weinstein, 2008), although this situation is ameliorated when the teacher belongs to the same ethnic group than the students (Dee, 2005). Other research has also found that schools that promote integration, positive norms, values and academic aspirations also have narrower ethnic achievement gaps (H. M. Marks, 2002; Rich, Ari, Amir, & Eliassy, 1996).

Researchers also warn that ability grouping, tiering or tracking are practices that worsen the ethnic achievement gap because minority students are generally not enrolled in the most advanced groups, which implies they do not get the chance to obtain top marks (Jackson, 2012; Strand, 2012). However, remedial ability grouping in early years has been shown to improve the academic achievement of students with language needs (Robinson, 2008) and that forcing ethnic integration may result in increased prejudice and avoidance between ethnic groups (McKeown & Dixon, 2017).

As discussed in section 3.3.1, tiering is not a practice in Colombia, as all students within the same school take the same courses (same subjects and same level) and there is a large set of compulsory subjects. However, schools do offer different types of focus (academic or vocational), which may lead to differences between students in different schools rather than differences within schools, as further explored in chapter 6. The data explored in this thesis does not include measures of school climate, norms or values, but these results may help to explain why some schools are found to have a narrower ethnic achievement gap in chapter 9, which explores possible interactions for the ethnic achievement gap and its components.

2.5.2 Ethnic Composition and Peer Effects

School ethnic composition has been a matter of debate in the US, given the dissolution of their formal segregation system in the 1950s. The proportion of Black students within schools is related to lower scores for Black students and, to a lower extent, for White students (Hanushek et al., 2009; Mickelson & Bottia, 2009). This is a regularity that occurs in many other countries, including Latin American countries (Canales & Webb, 2018; Fruehwirth, 2013; McEwan, 2003, 2004; McEwan & Trowbridge, 2007; Noe, Rodriguez Cabello, & Zuñiga, 2005; Verhaeghe, Vanlaar, Knipprath, De Fraine, & Van Damme, 2018).

There are three recurring explanations for this effect. First, that teachers have low expectations about certain ethnic groups and react by lowering their instruction level when facing a large proportion of students of these ethnic groups (Black in the US, Black Caribbean and Black African in England, and Māori and Pasifika students in New Zealand, for example) (Burgess & Greaves, 2013; Gershenson, Holt, & Papageorge, 2016; Sacerdote, 2011; Turner, Rubie-Davies, & Webber, 2015). There is qualitative evidence that this is also the case in Colombia (Valoyes Chavez, 2015), which may be an explanation for the large school contextual effects of ethnicity found in chapter 6, which are not fully explained by the student and school characteristics observed in chapter 8.

Second, the existence of pure peer effects, that result from students' interactions. Fordham and Ogbu (1986) formulated the "acting White" hypothesis, according to which Black students in the US think academic success is entitled to White students. Therefore, high achievement is perceived as "acting White" and thus betraying their ethnic community. Consequently, Black students pressure their Black peers to avoid excelling in school, and therefore, Black students do not fulfil their academic potential. Fryer and Torelli (2010) and Patacchini and Zenou (2016) tested this hypothesis in the US, finding that while popularity increased with grades for White students, the opposite was true for Black and Hispanic students and that this effect was more pronounced for boys than for girls and for state-school than private school students. Nonetheless, this hypothesis has met no support by other authors in the US, the Netherlands and Russia (Ainsworth-Darnell & Downey, 1998; Ivaniushina & Alexandrov, 2018; van Tubergen & van Gaans, 2016).

Finally, these effects are linked to the correlation between the school ethnic composition and factors that affect achievement. These include access to resources, such as the ability to retain better-qualified teachers, the influence of ethnic composition in school climate and relationships between students and teachers, which tend to favour academic achievement in schools with larger proportions of the ethnic majority (Agirdag, Demanet, Van Houtte, & Van Avermaet, 2011; Harris, 2010; Liu, Van Damme, Gielen, & Van Den Noortgate, 2015; Palardy et al., 2015; van Ewijk & Sleegers, 2010b). However, it is hard to find empirical support for a particular theory using observational data, as is the case of this thesis.

2.5.3 School Quality

Finally, school level explanations assert that minority students disproportionately attend schools with a lower quality, which hinders their academic achievement (Rosenbaum & DeLuca, 2008). This hypothesis has been tested under the umbrella of the 'opportunity to learn' concept, which refers to the ability of minority groups to access qualified teachers, adequate infrastructure, and pedagogical material (Cervini, 2011; Schmidt, Burroughs, Zoido, & Houang, 2015; Schmidt, Zoido, & Cogan, 2013). Chapter 4 shows that in Colombia, minority students tend to attend schools with characteristics that have been associated with low performance, which makes this a possible explanation for the ethnic achievement gap in the country.

Another measure of school quality is school effectiveness. Roughly, a school is effective if its students perform better than students in the other schools with a similar student intake (Chapman, Muijs, Reynolds, Sammons, & Teddlie, 2016; Leckie, Pillinger, Jenkins, & Rasbash, 2010; Willms, 1992). The hypothesis is then that schools are less effective in promoting learning for minority students than for students of the ethnic majority, taking place what is called differential school effectiveness (Kyriakides & Creemers, 2018). In England, differential school effects by ethnicity have been found at Year 11 (age 16) between 1985 and 1992 (Nuttall, Goldstein, Prosser, & Rasbash, 1989; S. Thomas et al., 1997) but not after 1995 for different stages (Leckie & Goldstein, 2019; Strand, 1999, 2010, 2014a, 2014b, 2016). Around the world, there is evidence that schools tend to be less effective for ethnic minorities (Biedinger, Becker, & Rohling, 2007; Chatterji, 2006), although in Latin American countries the focus has been on gender (e.g. Muñoz-Chereau, 2018) and not on ethnicity. A school-fixed-effects approach using Peruvian data found that schools have a 0.25 SD weaker effect on ethnic minorities than on White students (Glewwe, Krutikova, & Rolleston, 2017).

2.5.4 Students or Schools?

Given the competing explanations for the ethnic achievement gap at the student and school levels, researchers have attempted to separate the ethnic gap in parts that can be attributed to the student and the school levels. In the US, the decrease in the Black-White achievement gap at all ages between the 1970s and 1980s and the later increase in the 1980s triggered a debate about the extent to which these patterns could be attributed to students or schools. The methodological debate that this question originated, revisited in chapter 6, left extreme claims, with one side stating that schools were only responsible for 37% of the Black-White gap (Cook & Evans, 2000), while other side assuring that 87% was attributable to the schools (Hanushek & Rivkin, 2006).

As further discussed in chapter 6, Cook and Evans (2000)'s methodology has been replicated to study achievement gaps in England and Latin America, with within-school differences found to be more important in England (Dustmann et al., 2010) than in Guatemala, Bolivia and Chile (McEwan, 2004; McEwan & Trowbridge, 2007). Chapter 6 shows that in Colombia, differences between local authorities (LAs) are more important than differences between schools.

2.6 Higher-Level Explanations for Ethnic achievement Gaps

Other researchers have suggested that the ethnic achievement gap is not a problem of students or schools, but of behaviours that hinder ethnic minorities or of the way the school system and the society are organised. The analysis in this thesis does not allow distinguishing between these potential explanations. However, the results in chapters 6, 8 and 9 show that students and schools are not in total control of the ethnic achievement gap and higher-level explanations are required.

Critical race theorists view racism, and institutional racism in particular, as a feature of US and UK societies (Gillborn, 2006, 2008; Howard & Navarro, 2016), where they have focused their work. They argue that racism is incorporated into regular behaviour by which the ethnic majority exerts power over the minority groups, commonly through their exclusion. These collective and subtle actions are defined as racism from this perspective because the outcome is that minority students are disadvantaged by the actions and omissions of the majority group. Scholars have also highlighted exclusion and stereotyping as mechanisms through which racism operates in Colombia (O. Barbary, Ramírez, & Urrea, 2002; Chaparro, 2009; Valoyes-Chávez & Martin, 2016).

However, other authors discard the idea of discrimination and assert that the problem is that the structure of society does not allow ethnic minorities to take advantage of gains in access to education and income (B. Duncan & Duncan, 1968). For Rothstein (2004), for example, providing 'middle-class experiences' to disadvantaged ethnic groups, such as visits to parks and museums, travelling and reading at home, is a tool to combat socioeconomic and ethnic inequality.

Finally, Bourdieu and Passeron (1990) argued that the school system reproduces and perpetuates the existent structures of power, which privilege certain groups, including the ethnic majority. This theory has met with support. For example, Yeh (2015) argued that the rigid structure of the education system and lack of performance feedback generate low self-efficacy, which translates into a lack of engagement and low achievement. In Colombia, schools and the White-centric curriculum have been pointed out as mechanisms that perpetuate the dominance of the White majority (Gallo Restrepo, Meneses Copete, & Minotta Valencia, 2014; Hellebrandová, 2014; Ojeda Rosero & González Gómez, 2012; Quintero Ramírez, 2017).

In this thesis, LAs are included in the analysis as a higher-level that captures aspects outside the control of students and schools. In Colombia, this level might be seen to represent institutional practices that favour or hinder ethnic minorities, although this thesis does not provide evidence to test any of the theories reviewed here.

2.7 Framework for the Thesis

Rothstein (2004, p. 5) conceptualised the ethnic achievement gap as "a phenomenon of averages" that combines information about the student's ethnicity and SES, which at the same time reflect differences in the set of experiences that socioeconomically advantaged students get, but their disadvantaged peers do not. In this thesis, this conceptualisation is extended to understand that the ethnic achievement gap is a combination of comparisons between White and minority students in different schools and LAs, and with different characteristics, such as SES.

Therefore, the overall ethnic achievement gap is first understood as a multilevel phenomenon of averages that combines information about different levels of the education system: students, schools and LAs, which means that the overall achievement gap can be decomposed into its components at each of these levels, as further explained in chapter 6. Each of these components also reflects the characteristics that, if differentially distributed for White and minority students, are combined into the overall ethnic achievement gap. When these different characteristics are considered, and only White and minority students with the same set of observed characteristics are compared when computing the differences in test scores, a conditional (on these characteristics) ethnic achievement gap is obtained. The difference between the overall and conditional achievement gap is interpreted as the proportion of the gap that is explained by these characteristics, as shown in chapters 7 (focusing on SES) and 8.

Finally, the overall gap also combines information from different subgroups of students, schools and LAs (e.g. socioeconomically advantaged vs disadvantaged students, state vs private schools and LAs with a high vs low crime rate) that may have a different ethnic achievement gap than the pooled population. These potential subgroups are analysed in chapter 9.

This framework allows incorporating the explanations for the ethnic achievement gap that have been proposed in the literature summarised in this chapter, and that can be tested using the available data (described in chapter 4). It differs from the frameworks commonly used in the analysis of ethnic achievement gaps (if any) in that it focuses on the gap itself, instead of borrowing a framework that focuses predicting academic achievement, such as the educational production function (Hanushek, 1970) and the value-added models in the educational effectiveness framework (see Palardy, 2008; S. Thomas, Salim, Muñoz-Chereau, & Peng, 2012; Willms, 1992).

This understanding of the ethnic achievement gap requires adapting the existing methods or their interpretation to match with the framework. Therefore, as outlined in chapter 5, there is an important methodological aspect of this thesis, which uses a combination of single- and multilevel-modelling analysis while using mediation analysis as a tool to understand the achievement gaps.

The framework presented here could be extended to recognise that the overall ethnic achievement gap also incorporates differences at additional levels of the education system (e.g. classes within schools or countries) and the broader student context (e.g. families, neighbourhoods or towns). Nonetheless, these cannot be separated from the overall gap in the context of this thesis, given the available data. As presented here, however, this framework is useful to answer the research questions in this thesis and contribute to the ethnic achievement gap literature, which is summarised in the next section.

2.8 Summary

This chapter showed that ethnic achievement gaps are a regularity of many education systems around the world and provided an overview of the explanations for them, while each research chapter (6, 7, 8 and 9) provides a tailored literature review for their specific research questions and the context chapter (3) reviews the specific literature for Colombia. This overview not only informed the conceptual framework of this thesis but also identifies three important gaps in the literature.

First, the evidence is dominated by industrialised countries, especially from the US, with a minimal discussion about ethnic achievement gaps in non-industrialised countries. As further discussed in section 3.4, the evidence in Colombia is extremely limited, with only one study. Second, large-scale analysis of the gap has focused on students and schools, overlooking explanations at higher levels such as LAs or regions. Finally, there is little reflection and consensus about the most appropriate methods to study this phenomenon. This is the case on two specific areas addressed in this thesis: the debate on how to decompose the achievement gap into differences within and between schools and LAs (in chapter 6) and the operationalisation of SES (in chapter 7).

3 Colombia: An Overview

3.1 Introduction

This chapter describes the Colombian context to allow for a deeper understanding of the importance of exploring ethnic achievement gaps in this country and of the potential explanations for the patterns that are found throughout the research chapters of this thesis. The chapter starts by presenting the broader Colombian context in section 3.2. Section 3.3 focuses on the education system and the role of each of its levels (national authorities, local authorities (LAs) and schools), following the framework presented in section 2.7. This contextualisation then allows understanding the literature on educational research in Colombia, which section 3.4 reviews. Section 3.5 summarises the connections between this and other chapters of this thesis.

3.2 The Colombian Context

Colombia is a South American upper-middle-income country with a population of 42.8 million in 2018, of which 77.8% live in urban settings (DANE, 2018). Administratively, the country is divided in 32 departments (*departamentos*), which are themselves divided into a total of 1,122 municipalities (*municipios*). As further explained in section 3.3, some municipalities and departments are also LAs.

With 87 Indigenous groups and three different Afrocolombian ethnicities (Black, *Palenquera* and *Raizal*); 64 different Indigenous and two Afrocolombian tongues (DANE, 2007), Colombia is self-recognised as a multicultural country. Spanish is the official nation-wide language, but these tongues also have an official character in their respective communities. According to the last available census, in 2005 Indigenous groups made up 3.4% of the population, Afrocolombians represented 10.6% of the population and the Romani community accounted for 0.01% of Colombians (DANE, 2005). The majority group (84.1% of the population) is called White-mestizo[1], which is the result of a mixture of Indigenous and Afrocolombian groups, 19th-century migrants from the Middle-East and colonisers from Spain (Ministry of Education, 2010). As further discussed in the data chapter (section 4.5), minority groups tend to be concentrated in the periphery of the country, especially in the Pacific and Caribbean coasts shown in the map in Figure 3.1.

Sixteen-year-old Colombian students have participated four times in the Programme for International Student Assessment (PISA) test (2006, 2009, 2012 and 2015) and in all occasions, the performance of Colombian students has been poor, with an average score equivalent to three years less education than that of students in countries that belong to the Organisation for Economic Co-operation and Development (OECD) (OECD, 2016).

Besides an ethnic distribution that is linked to its colonial past and distinctively poor performance on international tests, another particularity of Colombia is its long-standing internal armed conflict. This conflict has the potential to affect academic achievement through direct impacts on students, infrastructure and the organisation of the school system on specific communities (Williams, 2004). Even more, since the intensity of the conflict is unevenly distributed across regions, it also has the potential of affecting the ethnic achievement gaps, as studied in chapters 8 and 9.

The armed conflict in Colombia has a political origin, with conservatives and liberals trying to seize control of the national government since 1946. In the modern conflict, guerrillas with a left-wing political orientation, including FARC and ELN, combat against the official army and right-wing paramilitary groups. Alongside the actors in the armed conflict, drug cartels and organised crime also engage in violence. All these actors have been involved in kidnapping, forced displacement, massacres, enforced disappearance and crimes against people and their property (Constitutional Court, 2013; Red Nacional de Información, 2018). The dynamics of the conflict are complex, with the appearance and disappearance of different armed actors and varied responses from the national government to each group, from military confrontation to peace negotiation (Fajardo M, 2015; Ugarriza & Craig, 2012). This thesis does not offer an in-depth analysis of the effect of the conflict on the ethnic achievement gap, but this is a feature of the Colombian context that is considered in chapters 8 and 9.

[1]For simplicity, this group is referred as 'White' in this thesis.

Source: Instituto Geográfico Agustín Codazzi - IGAC (2012).

Figure 3.1: Official political and physical map of Colombia. Departments' names are capitalised while their capitals' names are not. Dashed lines represent international borders and solid lines denote limits between departments. Bogota, the capital of Colombia is also the capital of the department called Cundinamarca.

An additional feature of Colombia is its market-oriented education system, with relatively high levels of autonomy for LAs and schools, as further described in the next section.

3.3 The Education System

The Colombian education system is regulated by the General Law of Education (Congress of Colombia, 1994a), which defines a hierarchical structure for governmental institutions in charge of Colombian education. The Ministry of Education (MoE) is in charge of the 94 local authorities (LAs) (*Entidades Territoriales Certificadas*) that oversee around 32,019 schools providing all levels of compulsory education around the country. Although the law establishes principles to be followed at the national level, it also gives autonomy to schools and LAs, which means that schools' and LAs' decisions have the potential to influence the ethnic achievement gap. The following sections describe the rules that are formulated at the national level and the aspects that can be changed by LAs and schools.

3.3.1 National-Level Directives

The MoE oversees the education system at the national level, including its structure, for which the MoE defines the stages and sets a national curriculum for compulsory subjects (Congress of Colombia, 1994a, art. 148). Pre-higher education schooling can be broadly divided into three stages. Students enter the education system for the first time at pre-school, when they are no more than five-years-old (at the beginning of the school year). They then progress through elementary and secondary education by satisfying the requirements for each grade (Presidency of the Republic of Colombia, 1994), which implies that around 2% of students in primary and secondary education have repeated at least one grade (UNESCO Institute for Statistics, 2012). The net enrolment rate for primary education is 90%, but it only reaches 79% for secondary education (The World Bank, 2016).

Table 3.1 shows the expected student's age at each of these levels, in comparison with the expected ages in England, a context that may be more familiar to the reader. Even though both systems include similar phases, they are divided in different ways, and students are expected to remain in the school system for a longer time in England. Additionally, students' achievement is assessed at several points during their educational process in both systems, although the age of assessment again differs.

In Colombia, mandatory subjects[2] are assessed by the Colombian National Assess-

[2]The mandatory subjects in elementary school are natural sciences and environmental education; social sciences, history, geography, political constitution and democracy; artistic and cultural education; ethics

Table 3.1: Comparison of educational stages in the England and Colombia

	England				Colombia		
Age	Phase	Key Stage	Grade	Assessment	Phase	Grade	Assessment
3	Primary	Early Years	0	EYFSP-Early Years Foundation Stage	Pre-School	Pre- nursery	
4			Reception			Nursery	
5						Transition	
6		KS1	1	Year 1 phonics screening check	Primary	1	
7			2			2	
8		KS2	3			3	SABER 3º
9			4			4	
10			5			5	SABER 5º
11			6	KS 2 tests	Secondary	6	
12	Secondary	KS3	7			7	
13			8			8	
14			9			9	SABER 9º
15		KS4	10			10	
16			11	GCSEs		11	SABER 11º
17	Further education	A-levels or vocational	12				
18			13	A-Level exams			

Sources: Department for Education (2014) for England and Congress of Colombia (1994a) for Colombia.

ment Institute (Icfes) at the end of 3rd, 5th, 9th and 11th grade (ages 8/9, 10/11, 14/15 and 16/17, respectively), using a set of tests designed to match the national curriculum (Congress of Colombia, 2009). These exams, known as SABER, are compulsory for all (around 9 million) students in the country, including those in about nine thousand private (fee-paying and self-managed) and ten thousand state (funded and managed by LAs) schools in all LAs around the country (Ministry of Education, 2016b). The SABER 11 exam, administered at the end of 11th grade has been applied since 1968, while the

and human values; physical education, recreation and sports; religious education; humanities, Spanish language and foreign languages; mathematics; technology and informatics. In middle school, the compulsory subjects are the same as in elementary school, plus economics, politics and philosophy (Congress of Colombia, 1994a, art. 23 and 30).

exams for 3rd, 5th and 9th grade were introduced relatively recently, adopting a yearly frequency since 2012, as further described in section 4.2.

The SABER 11 exam is a high-stakes exam, as its results are used as an admission criterion by many higher-education institutions. Since 2014, the exam has been used to allocate government-funded scholarships for higher education to socioeconomically disadvantaged students (Ministry of Education, 2015a; Velez Bustillo & Psacharopoulos, 2019). The school's mean score in the SABER 11 exam has also been used to regulate the fees that private schools can charge (Congress of Colombia, 1994a, Art. 202) and, since 2015 all the SABER exams are used to create an indicator of school quality as part of a national school accountability exercise called 'E-Day' (*Día E*) (Ministry of Education, 2015b).

3.3.1.1 Educational Policy

In Colombia, each four-year-long government sets out an investment plan, called the National Development Plan (NDP), outlining how the national budget will be executed during those four years. The NDP includes the budget allocation for educational policy, which has rarely addressed educational quality as a target (Barrera-Osorio, Maldonado, & Rodríguez, 2012; Piñeros Jiménez & Rodríguez, 1998; Rentería, 2009). Instead, it has historically focused on increasing enrolment rates in all educational levels and areas of the country. Policies have thus focused on the construction of school infrastructure, the creation of flexible programs that tried to reach over-aged population and students in rural areas and removing all school fees in state schools[3]. The results are positive in this area, but the expansion on the supply of educational service has historically overlooked concerns of quality (Bentaouet Kattan, Barrera-Osorio, Walter, & Taboada, 2008; Delgado Barrera, 2014; OECD, 2016). Concerns about quality have been addressed with the creation of a national curriculum and educational standards (*lineamientos currículares y estándares básicos de competencias*) and new regulations for hiring and promoting teachers in state schools in the early 2000s (Presidency of the Republic of Colombia, 2002b, 2002c), and the introduction of the 'E-Day' in 2015, which uses the results of the SABER exams to redesign schools' pedagogical strategies with the aim to improve their average test scores in following years (Ministry of Education, 2015b).

The most recent NDP, setting the investment projects to take place between 2018 and

[3] School fees are described in section 3.3.3.

2022 (approved in May 2019), aims to tackle socioeconomic, gender and ethnic inequalities in several areas. In education, the flagship program is the creation of new school infrastructure to support the expansion of a seven-hour day of schooling in all state schools. Most importantly, for the first time the NDP recognises that ethnic achievement gaps are a problem in Colombia (Congress of Colombia, 2019; DNP, 2018a). The next section describes how this and other governments have addressed educational policy for ethnic minorities.

3.3.1.2 Ethnoeducation: Educational Policy for Ethnic Minorities

The mobilisation of Indigenous and Afrocolombian communities during the 1970s and 1980s, to gain autonomy and recognition of the value of their cultural traditions, religious practices and beliefs, led to the recognition of Colombia as a multicultural country in the current Constitution (1991) and the inclusion of ethnoeducation in the general law of education (Castillo Guzmán, 2008; Congress of Colombia, 1994a; Rojas Curieux, 2010). This law defines ethnoeducation as the education for groups or communities with their own tongue, culture, traditions and administrative autonomy that, besides following the principles that guide conventional education, supports the development of the identity, culture, language, beliefs and production and social processes (Congress of Colombia, 1994a, art. 55-63). The implementation of this law has resulted in an umbrella of policies that fall under the term ethnoeducation.

There are two outcomes of this law that affect all ethnic groups: First, the official recognition of ethnoeducation programs, created by local teachers (with the support of their communities) according to the language, customs and traditions of their communities. Second, the creation of higher education programs and financial incentives for teachers to learn pedagogical strategies and course content to teach under an ethnoeducation program. In practice, this implies that 1.4% of schools in the country offer educational programs based on ethnoeducation, with teachers who have been trained with this specific purpose.

Nonetheless, the priorities of Afrocolombian, Indigenous and Romani communities about the ethnoeducation agenda have been different, which has resulted in different treatment for each of these groups. While all ethnic groups aimed to gain recognition of the value of their beliefs and practices (Ministry of Education, 2018a), Afrocolombian groups prioritised the recognition of their contributions to historical developments in the country throughout the incorporation of Afrocolombian studies as a compulsory subject in the national curriculum (Congress of Colombia, 1993). In turn, for almost 30 years Indigenous communities have been negotiating an agreement on the creation of an Indigenous education system with two outcomes: an allocation from the national budget to Indigenous communities in 2001 (Congress of Colombia, 2001), and a special transitional regime in 2014, recognising their autonomy to manage their schools and other institutions (Ministry of Inner Affairs, 2014). The new NDP aims to strengthen ethnoeducation programs around the country. For Afrocolombian communities, the NDP outlines investments in professional and masters' programs for teachers' and school-leaders' specialised training in ethnoeducation programs, the creation of new pedagogical material for class instruction and the incorporation of digital technology for ethnoeducation. For Indigenous communities, the investments aim to strengthen their administrative capacity so they can increase their autonomy and lead their own education system. Both Afrocolombian and Indigenous groups also proposed rethinking the pertinence and suitability of the SABER exams for the students in these communities. This is one way in which each minority group has a different experience from the same policy, which may help explain why the results of research chapters 6, 7, 8 and 9 vary for each ethnic minority.

Ethnoeducation has been criticised as a policy that ignores diversity, promotes segregation and the hegemony of western knowledge and culture, and fails to address the infrastructure and educational resources requirements to successfully implement this program (Caicedo & Castillo Guzmán, 2011; Calvo Población & García Bravo, 2013; Castillo Guzmán, 2008; Ferrero Botero, 2015; Rojas Curieux, 2010). Nonetheless, Blanco álvarez (2008) and Caicedo Ortíz (2011) highlight the flexibility of the program and its positive effects on self-recognition and empowering of Afrocolombian communities. Whether these benefits and drawbacks of ethnoeducation are linked to ethnic differences in academic achievement is explored in chapters 8 and 9.

3.3.2 The Role of Local Authorities

As mentioned in section 3.2, local authorities (LAs) are either municipalities or departments of the country. In municipalities with more than 100,000 inhabitants, the LA is the municipal secretariat of education, whereas in less populated municipalities -if they lack the necessary administrative capability- the LA is the departmental secretariat

(Congress of Colombia, 2001). Currently, the MoE oversees 94 LAs, although this number changed until 2010, when some municipalities finished the administrative processes to take control over the education system. There were 80 LAs in 2008 and 84 LAs in 2009. Between 2010 and 2013, 62 of the 94 LAs were municipalities, which means that 66% of the LAs only have one municipality. This organisation explains the relevance of the hierarchical structure of this thesis, that considers LAs instead of municipalities or departments.

LAs are responsible for planning, providing, managing and inspecting compulsory education within their jurisdictions, as well as for implementing policies that the MoE formulate (Congress of Colombia, 2001). LAs can also design their own policies if they are not in conflict with the national directives and if they have enough resources, which come from each LA fiscal efforts or the national transfers described in section 3.3.2.1. This situation leads to variation between LAs in the kind of policies they adopt and in how they execute the nation-wide policies, according to the availability of resources. For example, some LAs decide to increase the provision of schools by creating and managing new schools, while others may decide to hire private schools on a per-student basis to serve students, similarly, as charter schools in the US or traditional academies in England[4]. As another example, all fees for state-schools were removed in Bogotá in 2010, while this only became national policy for primary and secondary education in 2012 (Presidency of the Republic of Colombia, 2011; Redacción El Espectador, 2010). All these decisions have the potential to affect ethnic inequality both within and between LAs. Since the viability of such decisions depends on the availability of funding, the next section explores how funding decisions are made.

3.3.2.1 Funding for Schools

Chapters 8 and 9 study to what extent the ethnic achievement gaps are explained by and vary according to the availability of resources at the LA-level. This section summarises the nature of the resources that each LA receives from the central government.

Transfers from the central government fall into two broad categories, according to their destination: running costs (*funcionamiento*), which are destined to pay for teachers' and administrative workers' salaries (including social care) and schools' public utilities;

[4]The current data does not allow to differentiate between state-schools and academies. Both types of schools are recorded as state schools.

and quality (*calidad*), which are used for investment in infrastructure and educational resources (e.g. books, desks, computers) and cannot be used for recurring payments (Congress of Colombia, 2001, ch. IV). While running costs are paid to the 94 LAs, payments in the quality category are made directly to the 1,122 municipalities within LAs.

The allocation of funds for running costs was higher for more dispersed LAs, but since 2011 they are also higher for LAs with higher proportions of socioeconomically disadvantaged and (since 2012) ethnic minority students. Nonetheless, teachers' salaries are fixed by a nation-wide salary and grading structure that rewards teachers of Indigenous students in Indigenous territories (Presidency of the Republic of Colombia, 2012a, 2012b). Similarly, before 2011, resources for quality were adjusted according to the poverty conditions of each municipality, but after this year it also considers the LAs' performance and improvement in educational outcomes (grade repetition, dropping-out rates, which affect around 2% and 8% of students, respectively (UNESCO Institute for Statistics, 2012), and average SABER test scores). In practice, not all LAs with similar conditions receive the same amount of resources per pupil due to further adjustments according to each LA's needs. Therefore, there are multiple potential explanations for the relationship between the two kind of central government transfers and the ethnic achievement gap that are studied in chapters 8 and 9. These include the effects of differences in the transfers, the rules for the allocation and the underlying variables that determine the amount to transfer.

As studied in the next section, schools are autonomous in Colombia. As a mechanism to prevent schools from wrongdoing, LAs are also in charge of inspecting private schools, but there is no resource allocation from the central government for this task. The result is that inspection is closer in those LAs with more availability of their own resources, which is also studied in chapters 8 and 9. Nonetheless, this means that school-based decisions are also crucial for understanding ethnic achievement gaps in Colombia.

3.3.3 School Autonomy

Around 16% of students attend the nine thousand private schools in Colombia, which are self-managed fee-paying institutions inspected (but not controlled) by their respective LAs, using the SABER 11 exam results as the main evaluation criteria. The annual fees that private schools can charge depend on such inspection and vary widely, from less than the average monthly salary to 40 times this value (Congress of Colombia, 1994a; Ministry of Education, 2018b). Therefore, families with different levels of income can decide to send their children to private schools, but only families with high incomes can afford to send their children to the highest-achieving private schools. These schools can also set their own admission criteria.

In turn, about 84% of students attend the around ten thousand state schools in the country, which do not charge fees and are under the direct control of the LAs. Although state schools do not charge fees for enrolment, before 2011 students attending state schools in some LAs were still responsible for additional costs, such as their student identification cards or their grade reports. These costs are covered for elementary-school students since 2011, and for students in all education levels since 2012 (Presidency of the Republic of Colombia, 2011), in an attempt to increase enrolment from students with the lowest levels of income around the country. Students are assigned to their nearest state school, which is required to admit all students, regardless of their prior achievement (Ministry of Education, 2003).

Since the data chapter shows that minority students are less likely to attend private schools, these differences have the potential to explain the ethnic achievement gap, as explored in chapter 8. Nonetheless, there are multiple ways in which schools can differ, since schools are, by law, autonomous institutions (Congress of Colombia, 1994a). Therefore, schools can design their educational program (following or not the national curriculum) and decide over the school year, the school day, their pedagogical strategies and, in general, their strategy to impart education (Congress of Colombia, 1994a; Presidency of the Republic of Colombia, 2002b). This autonomy also represents a source of variation in school characteristics, which may explain part of the ethnic achievement gap and lead to between-school differences in the ethnic achievement gap, as chapters 8 and 9 explore.

Despite this freedom, schools follow some conventions and can be grouped according to their characteristics. There are mainly two types of school year: type A, from February to November; and type B, from September to June (Congress of Colombia, 1994a; Presidency of the Republic of Colombia, 2002a). The SABER 11 exam is carried out twice a year, once around March for school-year type B, and the other around August for school-year type A. For historical reasons, schools in LAs in the south-west of the country[5] run under a school-year type B, while schools in the rest of the country run under the school-year type A. Arguing difficulties with the national transfers' schedule and their ability to execute the assigned resources, the LAs in this region decided to switch to school-year type A, affecting all the state schools under their control, but not the private schools, which have the freedom to decide their school-year type (Colombia Aprende & Ministry of Education, n.d.; Navarro Piedrahita & Tovar, 2008; Secretaría de Educación Municipal de Santiago de Cali, 2008). After a short transition period, all students in state schools around the country sat the SABER 11 exam simultaneously for the first time in 2010 (Redacción de El País, 2010). After this change, all state schools and most private schools run under school-year type A. Private schools in the south-west and some schools on the Caribbean coast run under school-year type B. This change affected the comparability of the SABER 11 tests over time, as discussed in section 4.2.1. As discussed in chapter 6, this change in the school-year type may be reflected in the patterns for the ethnic achievement gaps.

Schools also decide on their school day, whether they provide classes during the morning, afternoon or the full day[6]. In theory, schools must impart 30 hours per week and 1,200 hours per year for secondary education (Presidency of the Republic of Colombia, 2002a), which would be the minimum to be considered a full day of school. However, as a strategy to increase enrolment, schools serve students during the morning or the afternoon, with varying lengths of the school day.

Finally, schools can decide on a different focus of their educational program. There are three broad types of foci: academic focus, preparing students to pursue further education; technical focus, instructing students to participate in the productive sector; or teaching focus, which includes pedagogical training for future school teachers (Congress of Colombia, 1994a). Some schools also decide to have a mixture between two or more of these foci, with the academic and technical combination being the most common. Another (non-mutually exclusive) option is that the whole curriculum is based on an ethnoeducation program, which can be bilingual and incorporate specific subjects of importance for a specific community. Despite this freedom of choice, at least 80% of the curricular plan of every school must include the compulsory subjects that are outlined in the national curriculum.

[5] An area with a high proportion of minority students, as shown in section 4.5.

[6] There are also schools offering programs during the night or over the weekend. However, these schools usually serve the adult population, who are not the focus of this thesis.

These school characteristics are observed in the SABER 11 data and described in section 4.4. They are analysed in chapters 8 and 9 of this thesis and in the educational research about Colombia described in the next section.

3.4 Literature Review for the Colombian Context

This section summarises two branches of educational research in Colombia that are relevant for this thesis, especially for chapters 8 and 9. First, section 3.4.1 summarises the findings of the dominant branch of quantitative educational research, which aims to find predictors of academic achievement. Then, section 3.4.2 summarises the research findings of quantitative and qualitative research on the role of ethnicity on academic achievement.

There are two reasons to present this context-specific review as part of the context chapter and not in the literature review (chapter 2). First, in Colombia, as in other countries, the problems that are studied in educational research and the potential explanations for them are very context-specific and therefore require some familiarity with the context to be understood. Second, chapter 2 focused on the ethnic achievement gap literature, but in Colombia, there is only one study in this area. Therefore, it was necessary to broaden the scope of the literature review for the Colombian context to support the model construction and interpretation of the substantive findings of the thesis.

3.4.1 What Variables Predict Achievement in Colombia?

Besides the work of Sánchez-Jabba (2011) and Bonilla Mejia (2011), summarised in section 3.4.2, ethnicity is not considered in the relatively large body of research in education in Colombia that has aimed at finding the best set of predictors of academic achievement. However, this literature provides a list of variables for which there is a strong consensus regarding their influence on academic achievement, which will be useful to explore potential explanations for the ethnic achievement gap in chapter 8 and defining potential subgroups of students, schools and LAs in chapter 9, following the methodological argument in chapter 5. Comparisons of the size of these effects are difficult because the studies use different methodologies and, when estimating the same kind of model, they usually include a different set of controls. However, the substantive findings are reasonably consistent, as further discussed in the following sections.

3.4.1.1 Student Characteristics

In Colombia, where longitudinal data at the student-level is not available, socio-economic status (SES), however measured[7], is usually the strongest student-level predictor of academic achievement, which is higher as SES increases (Correa, 2004; García Villegas, Espinosa Restrepo, Jiménez Angel, & Parra Heredia, 2013; Núñez, Steiner, Cadena, & Pardo, 2002; Rozo Alzate, 2017). Age and gender are also important predictors of achievement. Nonetheless, unlike what happens in the UK and other Western countries (Connolly, 2006; Crawford, 2007; Fleischmann et al., 2014), where girls and older students achieve higher test scores, in Colombia girls and older students obtain lower scores (Barrientos Marín, 2008; Barrientos Marín & Rios, 2007; de Jorge-Moreno et al., 2018; Gaviria & Barrientos Marín, 2001a; Martínez Barrera, 2012; Orjuela, 2013; Zambrano Jurado, 2013). The conventional explanation for this last result is that, in Colombia, age is a proxy for grade repetition and therefore older students are already more likely to obtain lower test scores. In contrast, in countries like the UK (with effectively no grade repetition and strict enrolment policies according to the month of birth) age is a proxy for exposure to the education system before testing (Crawford, 2007; Sutherland, Ilie, & Vignoles, 2015).

The household size (number of members in the household) is also negatively related to achievement. The explanation for this usually relates to resource competition regarding both, money and parental attention (García Villegas et al., 2013; Núñez et al., 2002). Achievement is also lower for students who work (instead of staying full-time in education) (Correa, 2004; Sánchez-Jabba, 2011; Soto Ceballos, 2014). Students who also work tend to attend school during the evenings or weekends and are forced to provide for themselves or their families and thus come from a disadvantaged background, which makes them more prone to get lower test scores.

As the data chapter (section 4.3) shows, minority students tend to be overaged and come from larger families with a more socioeconomically disadvantaged background. Therefore, all these variables are potential explanations of the achievement gap to be explored in chapter 8.

[7]For more details about how SES is measured in Colombia please refer to appendix A.1.1.

3.4.1.2 School Characteristics

In 1999, the National Bureau of Statistics (DANE) collected data about the infrastructure and educational resources around the country. These data were used in two ways: To estimate the efficiency of Colombian schools using a non-parametric technique called Data Envelopment Analysis[8] (Barrera-Osorio & Gaviria, 2003; Iregui, Melo, & Ramos, 2007) and to estimate the effects of these variables on students' test scores (Gaviria & Barrientos Marín, 2001b, 2001c; C. Rangel & Lleras, 2010).

In the first group, Barrera-Osorio and Gaviria (2003) and Iregui et al. (2007) in 1999 and 2002, respectively, estimated a positive relationship between students' SABER 11 test scores and educational resources. They found that schools could improve the way they use the resources they already have to produce better academic achievement outcomes. In the second group, using the 2003 SABER 11 data for Cartagena (a city and LA in the Colombian Caribbean coast), C. Rangel and Lleras (2010) found that educational resources (science equipment, audiovisual equipment, computer room, library, language laboratory, sport facilities, counsellors and teachers' education) and teacher-student ratio have a stronger positive effect on maths achievement than physical infrastructure (toilets, sinks, cafeteria, health and management staff). Gaviria and Barrientos Marín (2001b, 2001c) also found that teacher's education, the teacher-per student ratio, and the schools' educational resources (laboratories and libraries) have a positive effect on 11th-grade students' achievement in Bogota, even after controlling for student's characteristics. This variability implies that, as the international literature reviewed in chapter 2 suggests, differential school quality is a potential explanation for the ethnic achievement gap in Colombia, which is a possibility explored in chapter 9.

The data collected by DANE has not been updated, and therefore, studies referring to more recent achievement data are unable to use these school infrastructure and resources variables. Instead, the studies have used the school characteristics that are provided by the SABER 11 data. Studies usually find that students attending school for the full day and those in an academic (instead of a technical or teaching) programs perform better than students who attend schools during the night or the weekends (Abadía

[8] As Barrera-Osorio and Gaviria (2003) explained, this technique finds the highest average achievement a school with given characteristics could potentially achieve. Then, the efficiency of each school is estimated as its deviation from this optimum, given its characteristics. The closer to the optimum, the more efficient the school.

& Bernal, 2017; Alviar & Polanía, 1992; Barrientos Marín, 2008; Bentaouet Kattan et al., 2008; Correa, 2004; Gaviria & Barrientos Marín, 2001a; Iregui et al., 2007; Martínez Barrera, 2012; Piñeros Jiménez & Rodríguez, 1998; Rozo Alzate, 2017; Soto Ceballos, 2014). Chapter 8 shows that these variables also explain some of the ethnic achievement gaps.

Whether or not state schools perform better or worse is somewhat more controversial. On average, state schools perform worse than private schools (Barrera-Osorio et al., 2012). However, depending on the set of control variables and the method of analysis, the difference between private and state schools remains favouring private schools (Gaviria & Barrientos Marín, 2001b, 2001c; C. Rangel & Lleras, 2010; Vertel Morinson, Cepeda Coronado, & Lugo Hernández, 2014), the difference disappears (de Jorge-Moreno et al., 2018) or favours state schools (Bonilla Mejia, 2011). Chapter 8 shows that ethnic differences in enrolment in private schools partially explain the ethnic achievement gaps.

3.4.1.3 Administrative Division Characteristics

As explained in section 3.3.2, LAs are the relevant higher-level unit of analysis as they manage resources and policies that affect educational decisions in different regions around the country. Nonetheless, in the literature, authors usually focus on municipalities, an administrative unit that only in 6% of the cases is in control of educational decisions.

Loaiza Quintero and Hincapié Vélez (2016) analysed the geographical clustering patterns of the SABER 11 language and maths test scores. They found that municipalities that are surrounded by high-performance municipalities were more likely to keep achieving high test scores. However, they did not discuss whether this was the case because these municipalities belonged to the same LA or whether other mechanisms explained this pattern.

García Villegas et al. (2013) also focused on the effect of the municipality characteristics on students' academic achievement in 2011. Using the SABER 11 data, they found that economic development and the functioning of the judicial system were associated with higher achievement. In turn, national transfers for education had a negative effect on the student's achievement. They did not find a significant effect of the homicide rate, the internal displacement rate or the population size.

Similarly, Mina Calvo (2004) did not find an effect of conflict on the municipalities average achievement between 1996 and 1999, but did conclude that poverty and inequality at the municipality level had a negative effect on academic achievement. Besides, Barrera-Osorio and Gaviria (2003) found a negative effect of the municipality poverty on school effectiveness.

Since section 4.5 in the data chapter shows that minority students tend to live in places more heavily affected by the conflict, and with a more deficient management of resources, it is not surprising that chapter 8 finds that these variables are important explanations of the ethnic achievement gap, especially at the LA-level.

3.4.2 Ethnicity in Educational Research in Colombia

Research has shown that ethnic minorities are disadvantaged in terms of outcomes in the labour market (including likelihood of unemployment, salaries and job quality) and educational attainment (with lower enrolment rates and higher drop-out rates), and are also disproportionally affected by the armed conflict in the country (ILO, 2007; Manrique et al., 2003; Mora & Arcila, 2014; Pérez Marulanda & Mora, 2014; Romero Mora et al., 2017; Viáfara López & Serna Alvarado, 2015). Nonetheless, the relationship between ethnicity and academic achievement has not been widely explored.

Sánchez-Jabba (2011) is the first quantitative study to describe the ethnic achievement gap in Colombia. Through this thesis, there are multiple references to his work, focusing on the aspects of relevance in each of the research chapters. Sánchez-Jabba (2011) compared the SABER 11 maths and Spanish test scores of White and minority students, without differentiating among ethnic minority groups. He used the Oaxaca (1973) decomposition (described in section 5.3) and quantile regression to examine the extent to which the achievement gaps could be attributed to gender, zone (urban or rural), household size, mother's educational attainment, family income, whether the student has a part-time job, school-day type, school fees, and the administrative department's ethnic composition. Sánchez-Jabba (2011) performed this analysis at the national level and separately for each of the 32 departments in the country, using the 2010 SABER 11 test scores for students attending schools with school-year type A[9].

[9]Please note that this sample selection implies that Sánchez-Jabba (2011) excludes from his research all those students in private schools with school-year type B, which are located in areas with a high concentration of ethnic minorities.

Sánchez-Jabba (2011) found that 59.9% and 54.3% of the 0.35 standard deviations (SD) and 0.29 SD ethnic achievement gaps in maths and Spanish, respectively, were explained by observed variables. Nonetheless, in departments with a high proportion of minority students, the ethnic achievement gap was only marginally related to observed characteristics. He also found a wider achievement gap for those students in the top quantiles of the test score distribution, which could not be explained by differences in observed characteristics. Additionally, he found that mother's education was the most important predictor of academic achievement for both White and minority students and that the school fees had a stronger positive effect on achievement for minority than for White students. Although the ethnic achievement gap was narrower for Spanish than for maths, the overall results held for both subjects. These findings led Sánchez-Jabba (2011) to suggest that the provision of schooling of better quality, which he related to higher school fees, would be an appropriate strategy to narrow the ethnic achievement gap, while closer parental involvement could improve the academic achievement of all students.

This thesis extends Sánchez-Jabba (2011)'s work by considering different cohorts, differentiating among the major ethnic groups, focusing on LAs instead of departments as the appropriate administrative division for educational policy, proposing a multilevel approach for the analysis of the ethnic achievement gap, as further discussed in each research chapter.

Although Sánchez-Jabba (2011) is the only quantitative study that focused on the ethnic achievement gap in Colombia, Bonilla Mejia (2011) included dummy variables for Afrocolombian and Indigenous students as one of the many control variables in a model that aimed to estimate the impact of the school day type on educational achievement. Using the 2009 SABER 11 database, Bonilla Mejia (2011) estimated an overall conditional[10] achievement gap of 4.1% and 0.7% of the White students score for Afrocolombian and Indigenous students, respectively, but did not discuss these findings any further.

Besides the studies evaluating the benefits and problems ethnoeducation, summarised in section 3.3.1.2, educational research on ethnicity has explored the role of schools and

[10]The control variables in Bonilla Mejia (2011) included: age, gender, zone, household size, mother's education, family income, whether the student works, school fees, school type, municipality population, fiscal performance and homicide rate and the municipality composition in terms of urban population, poverty and adult illiteracy.

universities in racism. In general, racism is described as a systematic problem originated by the colonial institutions that created a socioeconomic hierarchy based on ethnic groups, and that persists today (Chóez-Ortega, 2014; Hellebrandová, 2014). Schools and universities are seen as institutions that reproduce this structure that privileges White students, although, it is argued, racism is evidenced through processes of exclusion as opposed to confrontation or violence (Gallo Restrepo et al., 2014; Ramírez, Urueña, & Moreno, 2012).

3.5 Summary

This chapter presented an overview of the Colombian context, focusing on its education system, and summarised the relevant findings of educational research in Colombia. This presentation reinforces the relevance of this thesis by showing the need for evidence about ethnic achievement gaps that informs potential policy efforts to tackle ethnic inequality in educational achievement around the country.

This chapter also feeds the methodological decisions and the interpretation of results for the research chapters. In particular, the hierarchical structure of the school system, with schools controlled by LAs, and the different historical origins and policies for the ethnic groups make it necessary to consider these differences when decomposing the ethnic achievement gaps in chapter 6, but also when evaluating potential explanations for the gap, in chapters 7 and 8, and examining different subgroups (e.g. boys and girls, private and state schools) in chapter 9.

These potential explanations and subgroups are also linked to the context studied in this chapter. School characteristics such as the day-type, focus or an ethnoeducational component of the program, and LA characteristics such as the prevalence of violence are specific to the Colombian context and are not considered in the international literature reviewed in chapter 2. The next chapter complements this presentation of the context by describing the data for these characteristics.

4 Data

4.1 Introduction

This chapter builds on the description of the Colombian context in chapter 3 by introducing the SABER 11 database and presenting descriptive statistics for student, school and local authority (LA) characteristics. The description of these data informs the methodological approach (chapter 5) and the interpretation of the findings in chapters 6 to 9.

Section 4.2 justifies the selection of the database. Sections 4.3, 4.4 and 4.5 describe students, schools and LAs in the dataset, focusing on ethnic disparities. Section 4.6 describes the limitations of the data and section 4.7 summarises.

4.2 Data Sources

Colombia has participated in several international and regional assessment ssuch as the Programme for International Student Assessment (PISA) (2006, 2009, 2012 and 2015), Trends in International Mathematics and Science Study (TIMSS) (1995 and 2007), Progress in International Reading Literacy Study (PIRLS) (2001) and the first (1997), second (2006) and third (2013) versions of the Regional Comparative and Explanatory Study from the Latin American Laboratory for Assessment of the Quality of Education (LLECE)–PERCE, SERCE and TERCE, by their acronyms in Spanish. These are potential databases to address the research questions of this thesis (section 1.2). Nonetheless, they do not identify ethnic minorities. PISA, TIMSS and PIRLS identify immigrant students and those whose first language is not the national language (National Center for Education Statistics, 2016), and studies from the LLECE identify Indigenous and non-Indigenous students, but not other ethnic minorities such as Afrocolombians (Atorresi et al., 2010; Casassus et al., 2001; UNESCO-OREALC, 2016).

As mentioned in section 3.3.1, the Colombian National Assessment Institute (Icfes) administers nationwide compulsory exams that take place in 3rd, 5th, 9th and 11th grade (ages 8/9, 10/11, 14/15 and 16/17, respectively), providing censuses of students in these stages. The SABER 11 exam has been applied since 1968 and applied twice a year for at least the last 20 years, while the exams for 3rd, 5th and 9th grade adopted a yearly frequency since 2012 and students cannot be followed over time since there is no unique identifying code for each student. Furthermore, student-level ethnic information is not available for the 3rd, 5th and 9th-grade databases.

In turn, the SABER 11 database includes student-level information on a rich set of demographic (including ethnicity) and socioeconomic variables and the test scores for 11th-grade students. These are available from 2000, when SABER 11 began to be a competency-based exam. However, information about students' ethnicity began to be collected in 2005. Additionally, the exam has gone through changes during this time, as discussed in the next section.

4.2.1 The SABER 11 Test over Time

Icfes has experimented with the structure of the SABER 11 in an attempt to improve the quality of the test (Icfes, 2017). This implies that there have been four different versions of the competence-based SABER 11 exam[1]: 2000-1 to 2007-1, 2007-2 to 2009-2, 2010-1 to 2014-1 and 2014-2 until the present (Icfes, 2013, 2015).

Between 2000-1 and 2014-1, the SABER 11 exam consisted of multiple-choice questions divided into a compulsory core (maths, Spanish, biology, chemistry, physics, history, geography and philosophy) and a flexible component. The test scores, reported on a scale with a mean of 50 and a standard deviation of 10 in 2005, were computed using item-response theory to consider the difficulty of each question. There was no intuitive interpretation for these scores and their comparability between intakes within the same year and over time was enhanced using anchored items (Icfes, 2017). Cuevas Mendoza (2013) reported that the questions in the maths test do not favour any particular ethnic group. The total number of questions were reduced for the core part of the exam in 2007-2 (Icfes, 2013) and for the flexible component in 2010-1 (Icfes, 2015). In contrast, the last version of the exam drastically departs from the previous exams. The new exam comprises five areas (maths, critical reading, social sciences and citizenship, natural sci-

[1]These versions are labelled Year - Semester; 2000-1 is the first semester of 2000.

ences and English), includes multiple-choice and open-answer questions, and the scores on individual tests are norm-referenced with mean 50 and standard deviation 10 with respect to the 2014-2 exam, ranging between 0 and 100 (Icfes, 2015).

These changes imply that the longest period for which the core part of the exam was stable for students in both school-year types (A and B)[2] is 2008 to 2013. Assuring that the exam is the same for both school-year types is crucial, given the differences in the ethnic distribution of schools under both types of school-year. Besides, direct comparison of the raw test scores before and after 2010 is problematic, as a consequence of the change of state schools under school-year type B to school-year type A described in section 3.3.3[3].

Subsequently, the data used in this thesis corresponds to the cohorts between 2008 and 2013. The test scores have been normal-score transformed to follow an approximately standard normal distribution in each year. This transformation is a common practice in educational research (e.g. Leckie, 2009; Strand, 2016) to avoid violating the assumption that multilevel models' residuals follow a normal distribution. This decision implies that changes in the mean test scores between cohorts cannot be examined. Nonetheless, it is possible to compute differences between students taking the exam during the first and second semester of the same year. Furthermore, the ethnic achievement gap can be compared over time, under the assumption that the changes in the exam from one year to the next have the same on all ethnic groups.

The analysis in this thesis focuses on the SABER 11 maths test scores. This competency-based exam assesses the ability of students to use their maths knowledge in contexts of daily life, scientific reasoning and other subjects. The test covers three knowledge areas: arithmetic, algebra and elementary calculus; geometry; and probability and statistics. The assessed competencies are reasoning, communication and problem-solving. The test comprises of 24 multiple-choice questions with single and multiple answers. Although there is a time limit for the SABER 11 exam, there is no specified time for each test. Therefore, students can choose how to allocate their time to each subject (Icfes, 2010b).

As discussed in section 4.6, the database imposes restrictions on the kind of analysis that can be performed. Nonetheless, the data provide nation-wide census information

[2]Please refer to section 3.3.3 for a description of school-year types.

[3]Appendix A.2.3 offers more details about how this change affected the test scores reported by Icfes.

on the maths competencies of all 11th-grade (age 16/17) students in the country that allows answering the research questions in this thesis.

4.2.2 School and Local Authority Data Sources

The final dataset used through this thesis is the result of combining the Icfes' SABER 11 dataset for the 2008 to 2013 cohorts (which provides the test scores and self-reported student-level data) with school characteristics in the Ministry of Education's 2016[4] *Buscando Colegio* (BC, Searching for School) database (Ministry of Education, 2016a) and LA characteristics from two different sources: the Colombian Department for National Planning (DNP) for information about funding for education (DNP, 2014, 2018b) and the National Institute for the Victims (*Unidad para las Víctimas*) for information about violence (Red Nacional de Información, 2018). Joining these different data sources allowed for a large number of potential factors that could be related with the ethnic achievement gap, as explored in chapters 8 and 9.

After the data cleaning process described in appendix A.2.1 and A.2.2, the final database that is used for all research chapters includes regular students from 6 cohorts. This excludes 20.4% of students and 14.8% of schools in the original database, who are in adult education, a program for gifted or physically or mentally impaired students, or have missing data for any of the variables that are analysed through the thesis. The aim is to improve comparability both with the international literature and of the models that are estimated in chapters 6, 7, 8 and 9. Schools with less than five students have also been excluded to avoid including unreliable data into the school-level estimates (Maas & Hox, 2005).

As Table 4.1 shows, the number of students in each cohort and the number of schools have increased over time. This trend is the result of (i) the government efforts to increase student enrolment and provide more educational infrastructure, as discussed in section 3.3.1.1, and, (ii) changes in the data quality, which improve the chances of identifying a larger number of schools, as explained in appendix A.2.1. The number of LAs has also increased over time, as explained in section 3.3.2. There are data for all existing LAs in each year.

[4]Although the ideal would be to match the information of schools in the same year their students take the exam, there are no available earlier versions of this dataset.

Table 4.1: Number of students, schools and LAs in the database per year

Level	2008	2009	2010	2011	2012	2013	Total
LAs	80	84	94	94	94	94	94
↑							
Schools	7,143	7,477	7,706	8,039	8,261	8,384	9,080
↑							
Students	387,773	427,576	460,107	458,947	448,068	470,210	2,652,681

4.3 Students

The SABER 11 dataset reports the student's ethnicity in 19 categories, which respond to the students' ethnic identification and not to their 'objective' ancestry[5]. Given the complexity of the models that are estimated in chapters 8 and 9, these categories had to be collapsed into broader groups to allow for a larger number of students within each group. The ethnic groups that are analysed throughout the thesis are: White (not belonging to an ethnic minority, representing 92.4% of the students from all cohorts), Afrocolombian[6] (3.5%), Indigenous[7] (1.8%) and other minorities (2.3%). This last group includes Romani students (0.02% of all students) and students who consider themselves part of a minority group that is not among the options that Icfes provides[8]. Students with a mixed ethnicity may either select the 'other' category, one of the minority groups or may not consider themselves part of any minority group and therefore be classified as White.

Figure 4.1 shows that the sample's ethnic composition is stable over time, excepting for 2013, when there is a higher proportion of minority students. In this year, the sample composition may have been affected by a national strike that took place between August and September, causing logistic difficulties in the SABER 11 exam application in regions with a low proportion of minority students (El Espectador, 2013). Thus this strike may

[5]The literature has shown that ethnic self-identification and ancestry are highly correlated, but self-identification is more highly correlated with academic achievement (e.g. Herman, 2009).

[6]This category includes Black, Palenque (maroon) and *raizal* communities.

[7]The (major) Indigenous groups identified in the SABER 11 database are: Paez, Sikuani, Arhuaco, Embera, Guambiano, Pijao, Wayuu, Zenu, Pasto, Cancuamo, Inga, Tucano, Huitoto, and Cubeo (Icfes, 2013).

[8]These two categories are combined given the low number of Romani students and that their observable characteristics are similar to the ones of the group of other minority students.

have disproportionally prevented White students from taking the exam. Another possibility is that students who would normally pick that they do not belong to any ethnic minority (for example, mixed-ethnicity students) decided to choose one of the minority groups. This change in the ethnic distribution of the sample is linked to different results for the 2013 cohort in chapters 6, 8 and 9.

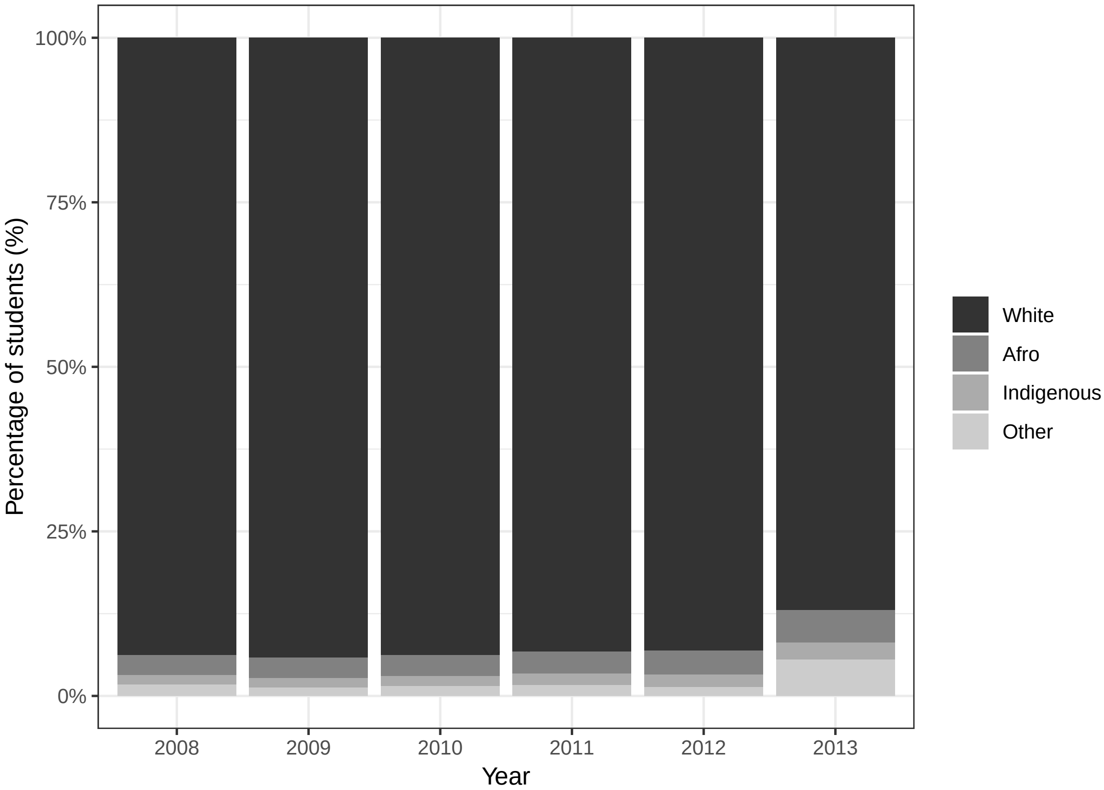

Figure 4.1: Sample's ethnic composition by year

As with other variables that are analysed through this chapter, ethnic differences in the distribution of maths test scores are stable over time. Therefore, the analysis for this chapter focuses mainly on the 2008 cohort, unless there are differences worth noticing. The same cohort is examined in detail in chapters 7 and 9.

Figure 4.2 shows the distribution of maths test scores for students of the four ethnic groups for the 2008 cohort, using density plots and boxplots. White students tend to outperform minority students with a mean test score 0.02 standard deviations (SD) above the overall average. In contrast, Afrocolombian and Indigenous students score on average 0.45 SD and 0.35 SD below the overall mean. The distribution of maths scores for the group of other minority students is very similar to that of the White students, with a mean score of 0.07 SD below the overall average.

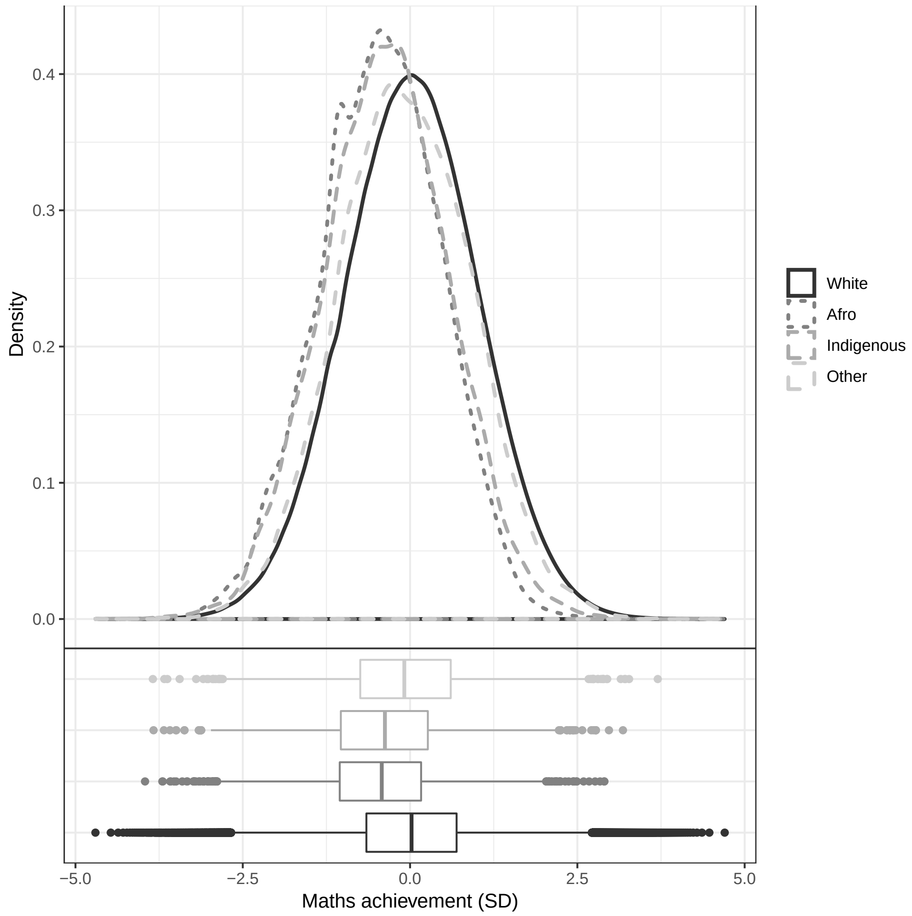

Figure 4.2: Distribution of the maths test scores by ethnicity in 2008

Figure 4.2 also shows that differences in achievement between ethnic groups are more pronounced for the top quartiles of their distributions. This mirrors to the finding in chapter 9 that those subgroups of students with characteristics that are linked with higher achievement are also the subgroups with wider ethnic achievement gaps. The figure also shows that there are substantial differences between students of the same ethnic group, which are also considered in chapter 9.

4.3.1 Socio-Economic Status

As discussed in sections 2.4 and 3.4, socio-economic status (SES) is considered one of the most important predictors of achievement in Colombia and one of the most important explanations for the existence of ethnic/racial achievement gaps around the world. In this thesis, SES is measured as the combination of parental education and occupation, and family income. While chapter 7 focuses on how to combine these variables to obtain a SES index, this section describes these variables as they are available in the SABER 11 dataset.

Figure 4.3 shows the distribution of the mother's and father's educational attainment. It is categorised according to the completion or not of each of the education levels described in section 3.3.1. There are three patterns worth noticing from this figure. First, although half of the parents of all ethnic groups completed at least primary education, 20.5% and 20.1% of White students have mothers and fathers with educational attainment above secondary education, while only 14.3% and 13.6% of Afrocolombian, 8.7% and 8.1% of Indigenous, and 21.2% and 21.6% of other minority students have mothers and fathers with this educational attainment, respectively. Second, as these numbers imply, mothers of all ethnic groups tend to have a higher educational attainment than fathers. Finally, for all ethnic groups, the proportions of parents with complete degrees (secondary, vocational or professional) are larger than the proportion of parents with incomplete degrees. This implies that when transforming these variables into years of education in chapter 7, the distribution will have some peaks at the years of education that imply completion of an educational level.

Differences in occupational status between mothers and fathers are more noticeable. As shown in Figure 4.4, most mothers of all ethnic groups are stay-at-home parents; with 56.9%, 58%, 71.4% and 55.2% of the mothers of White, Afrocolombian, Indigenous and other minority students, respectively, having this occupation. In turn, most fathers (35.6%, 47.2%, 58.2% and 34.9%, for White, Afrocolombian, Indigenous and other minority students, respectively) work as self-employed workers. This is consistent with high employment informality rates in Colombia, which is more prevalent among ethnic minorities (Pérez Marulanda & Mora, 2014; Romero Mora et al., 2017). Besides, both parents of ethnic minority students are less likely to have managerial positions. This implies that, for all ethnic groups, the main source of income in the family is generally the father's work, which tends to have a lower status for minority students.

Figure 4.5 shows that minority students tend to belong to families with lower income, with 44.3%, 65.2% and 31.9% of Afrocolombian, Indigenous and other minority students living in families with an income below the monthly minimum wage (MW)[9], while only

[9]The nation-wide minimum wage is fixed every year by the government to cover workers' basic needs, following a negotiation between unions and different industry representatives. For the purpose of this the-

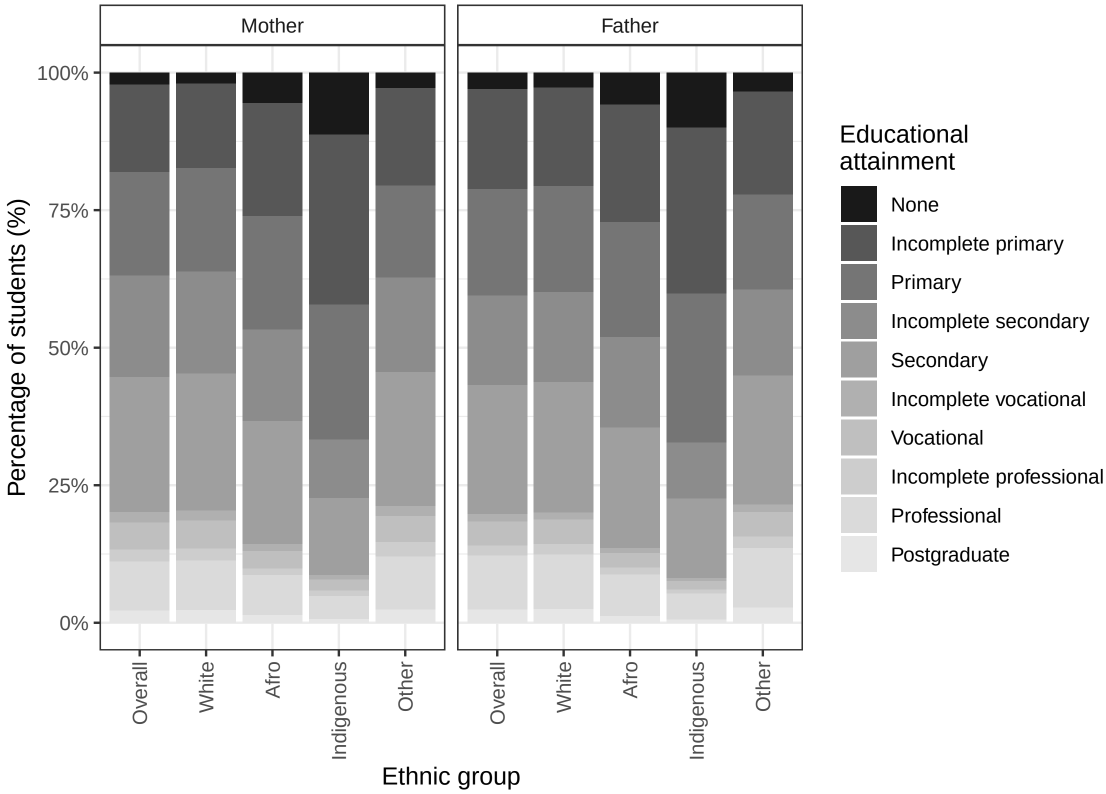

Figure 4.3: Distribution of parental educational attainment by ethnicity in 2008

25.4% of Whites doing so.

These figures show that, although students from ethnic minorities tend to belong to more socioeconomically disadvantaged families, in comparison with White students, there are also differences among minority groups, with Afrocolombian and Indigenous students being more disadvantaged than other minority students.

Three additional student variables are analysed in chapters 8 and 9: gender, age and household size. The mean and standard deviations for these last two and the proportion of girls for each ethnic group are shown in Table 4.2. As shown in this table, Afrocolombian and Indigenous students are older, on average than their White and other minority peers, although the average age for Afrocolombian and Indigenous students decreases over time, as shown in appendix A.2.4. Minority students also come from larger households, especially Indigenous students, than their White peers. Finally, Indigenous students are the only ethnic group with a lower proportion of girls than boys.

sis, it provides a unit of measurement that allows for comparisons over time, as it incorporates differences in the families' purchasing power from one year to the other.

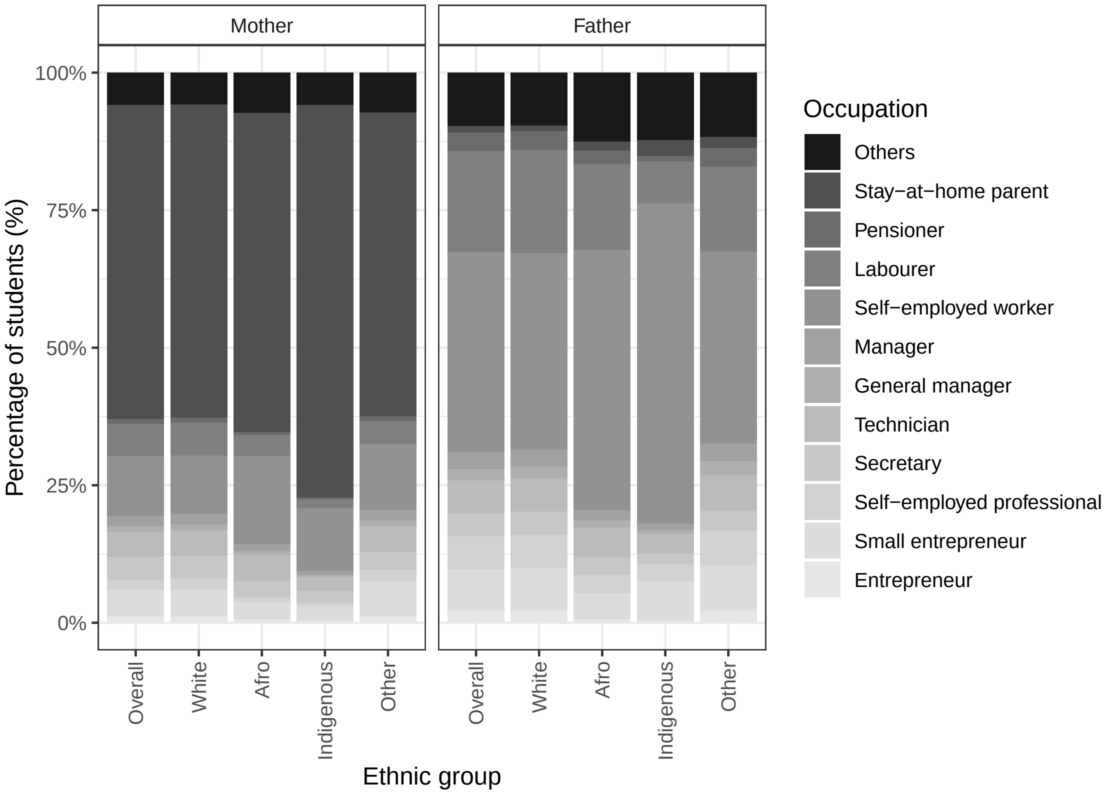

Figure 4.4: Distribution of parental occupation by ethnicity in 2008

4.4 Schools

As discussed in section 3.3.3, schools in Colombia can choose their characteristics, which may be related to the large variation in school mean test scores and therefore the ethnic achievement gap, as further explored in chapters 8 and 9. In the dataset, these characteristics include school type (state or private), school zone (urban, rural or both), whether the school's pedagogical program is based on ethnoeducation, their day type (morning, afternoon or full-day) and focus (academic, technical, teaching or academic and technical)[10].

There are noticeable differences in the school characteristics of private and state schools, as shown in Table 4.3 for the 7,143 schools for the 2008 cohort. Private schools are more likely to be in urban areas, offer full-day programs and an academic focus with

[10]There were some inconsistencies in the students' reports of the school-day type and the school focus. If this was the case, the school was assigned the value that most of the students selected. For example, if 90% of the students reported that the school-day type was full-day, the school was assigned full-day as the value for school-day type.

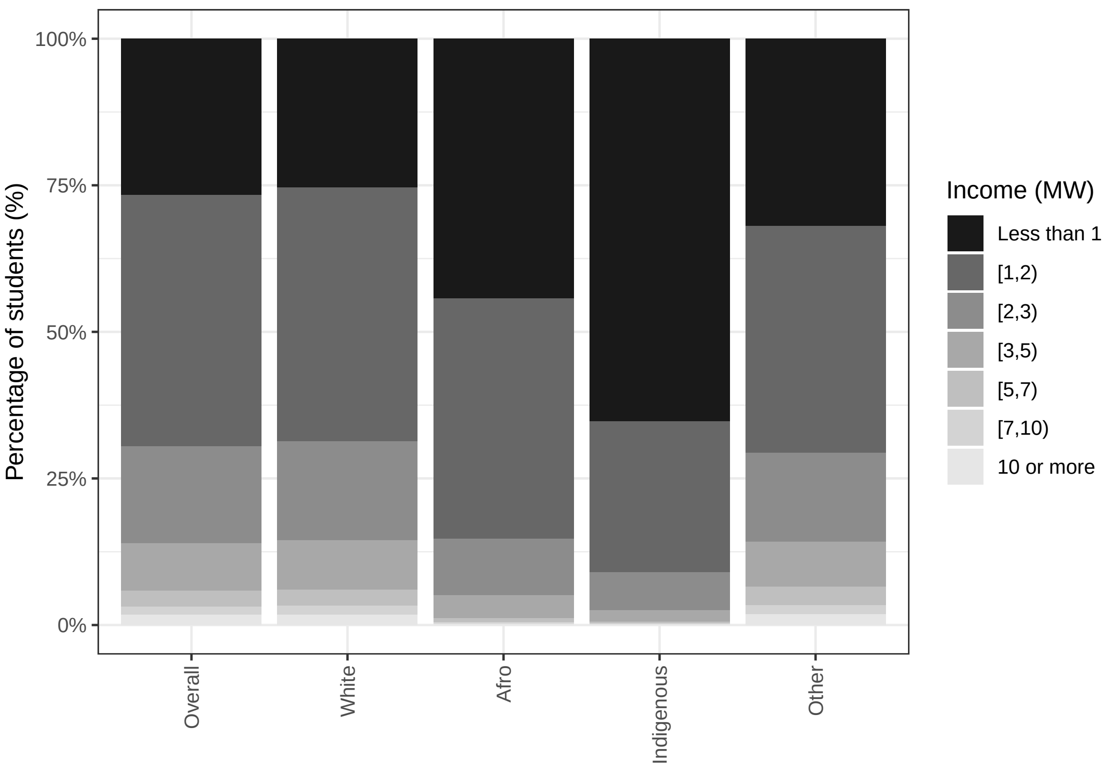

Figure 4.5: Distribution of family income by ethnicity in 2008

a program that is not based on ethnoeducation, in comparison to state schools.

This differentiation between state and private schools implies that the 24.2% of students who attend private schools are likely to attend very different schools than the rest of the population. What is more, only 7.4% of Afrocolombian and 4% of Indigenous students attend private schools, in contrast with the 25.1% of White and 21.8% of other minority students who attend these schools. These participation rates imply that state schools have an average proportion of minority students that is three times that of the private schools.

The analysis in chapters 6 to 9 includes the school proportion of each minority group as a strategy to decompose the ethnic achievement gap into its components at different levels of the education system (student, schools and LAs), as further explained in chapters 5 and 6. Therefore, the relationship between schools' ethnic composition and their mean test score is crucial through this thesis. Figure 4.6 represents this relationship for the school proportion of each ethnic group, whose distribution is shown by the rug plot, representing the distribution of the school proportion of students of each ethnic

Table 4.2: Descriptive statistics for age, household size and gender by ethnic group in 2008

variable	Overall	White	Afro	Indigenous	Other
Age	17.79	17.76	18.5	18.37	17.78
(years)	(2.8)	(2.79)	(3.2)	(2.58)	(2.45)
Household size	4.98	4.95	5.49	5.81	5.15
(members)	(1.86)	(1.84)	(2.14)	(2.21)	(1.98)
Female (%)	54.54	54.65	54.71	49.28	52.77

group, at the bottom of each graph. As shown in the figure, the distribution of the ethnic composition is skewed, with 65.7% of schools only serving White students, but also 1.2%, 0.5% and 0.3% of schools only serving Afrocolombian, Indigenous and other minority students, respectively. While the mean test score of schools mainly serving White students is around the national average, schools with a proportion of Afrocolombian, Indigenous and other minority students above 75% score on average 0.57 SD, 0.49 SD and 0.3 SD below the national mean.

As shown by different shades in Figure 4.6, these differences are linked to the school characteristics that change between state and private schools[11]. Only 0.6%, 0.1% and 0.7% of private schools have more than 75% Afrocolombian, Indigenous and other minority students, respectively, and 62.7% of the schools that score above the national mean are private.

One concern is that schools with very high proportions of minority students are a few small schools that appear in the sample only by chance. For this reason, Figure 4.6 also represents the number of students in the SABER 11 dataset for each school (a proxy for school size) by circles of different size. Schools that exclusively serve one of the ethnic groups tend to be smaller, with a median size of 28, 14, 21 and 11 students for schools only serving White, Afrocolombian, Indigenous and other minority students, respectively, in comparison to the median school of 34 students[12]. However, this pattern of small schools and schools with a high proportion of students of an ethnic group is consistent over time, as shown in appendix A.2.4.

[11] Differences in school ethnic composition according to other school characteristics are presented in appendix A.2.4.

[12] As mentioned in section 4.2.2, the sample has been restricted to schools with more than five students.

Table 4.3: Percentage (%) of state and private schools according to their characteristics in 2008

Characteristic	State	Private	Total
Zone			
Urban	47.75	92.82	63.29
Rural	31.33	7.14	22.99
Both	20.92	0.04	13.72
Day type			
Full day	17.76	54.79	30.53
Morning	68.07	38.31	57.80
Afternoon	14.17	6.90	11.66
Focus			
Academic	54.90	72.77	61.07
Academic and Technical	16.16	10.43	14.18
Teaching	6.41	6.57	6.47
Technical	22.53	10.23	18.28
Ethnoeducation			
Yes	2.01	0.28	1.41
No	97.99	99.72	98.59
Total	65.50	34.50	

4.5 Local Authorities

While White students live in all LAs in the country, minority students are located in the periphery of the country. As shown in Figure 4.7, 4.5% and 24.1% of students living in LAs near the Caribbean sea and the Pacific Ocean, respectively, are Afrocolombian, while 8% and 2.4% are Indigenous. Conversely, 86.2% of Afrocolombian and 80.6% of Indigenous students students live in these regions. Other minority students are more dispersed across the country, although only 13.9% of them live in the centre and south-east of the country.

Therefore, the ethnic composition of schools described in section 4.4 partially reflects the ethnic distribution of the LA, but may also reflect further selection processes in the allocation of students to schools. As with the school ethnic composition, the LA ethnic composition is used to analyse the components of the ethnic achievement gap at each level of the education system in chapters 6, 7, 8 and 9. However, this is not the

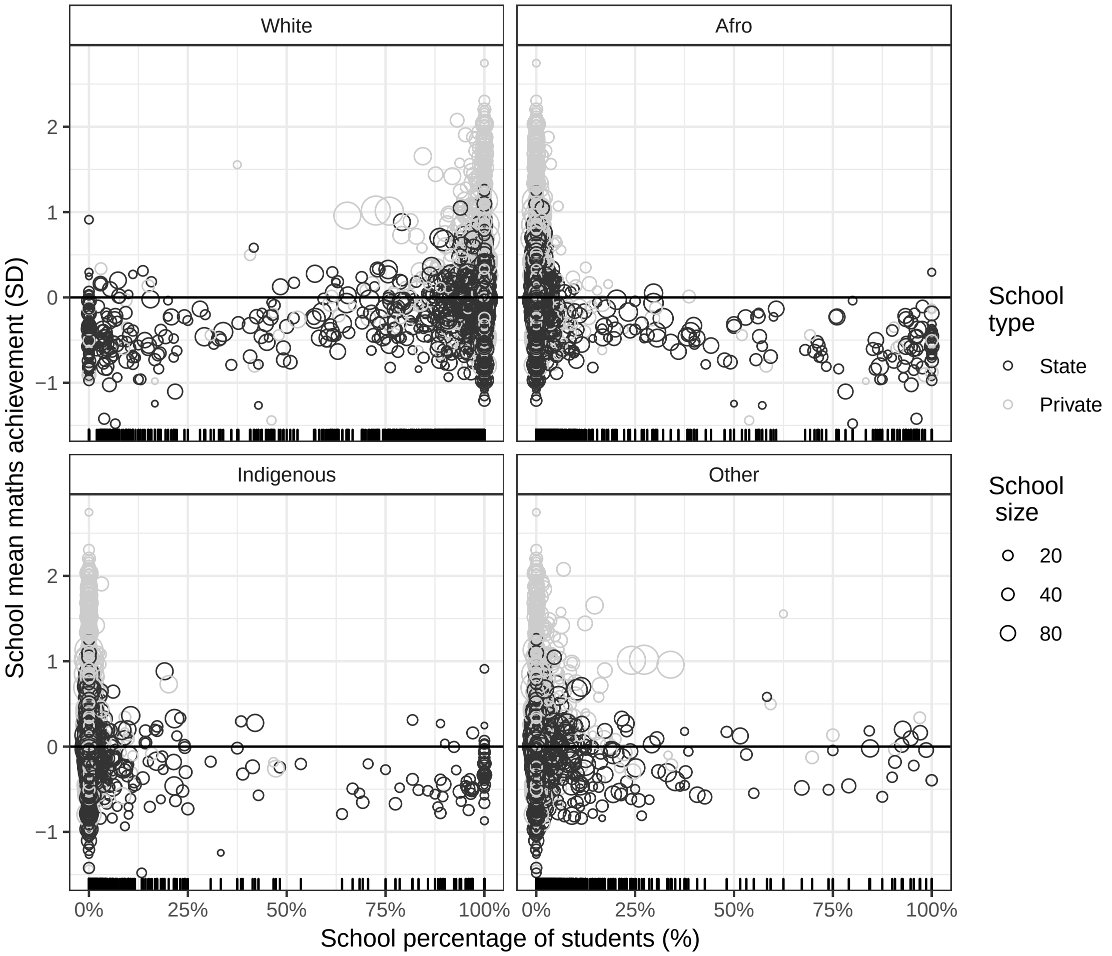

Figure 4.6: Relationship between school mean test scores and school proportion of ethnic group by school size in 2008

only variable of interest for the research chapters in this thesis. The following sections describe in more detail other LA characteristics.

4.5.1 National Transfers and Fiscal performance

Section 3.3.2.1 explained that there are two types of national transfers to fund education in each LA. Transfers for running costs and transfers for quality. Both kinds of transfers are measured on a per capita basis and in constant Colombian pesos ($) of 2007[13].

These resource-allocation rules create some variation between LAs in the availability of resources, with transfers from running costs ranging between $ 107,406 and $ 682,206 per student and transfers for quality being between $ 5,548 and $ 42,435. However, LAs

[13]The quality transfers were calculated as the sum of transfers for all municipalities in the LA over the LA population.

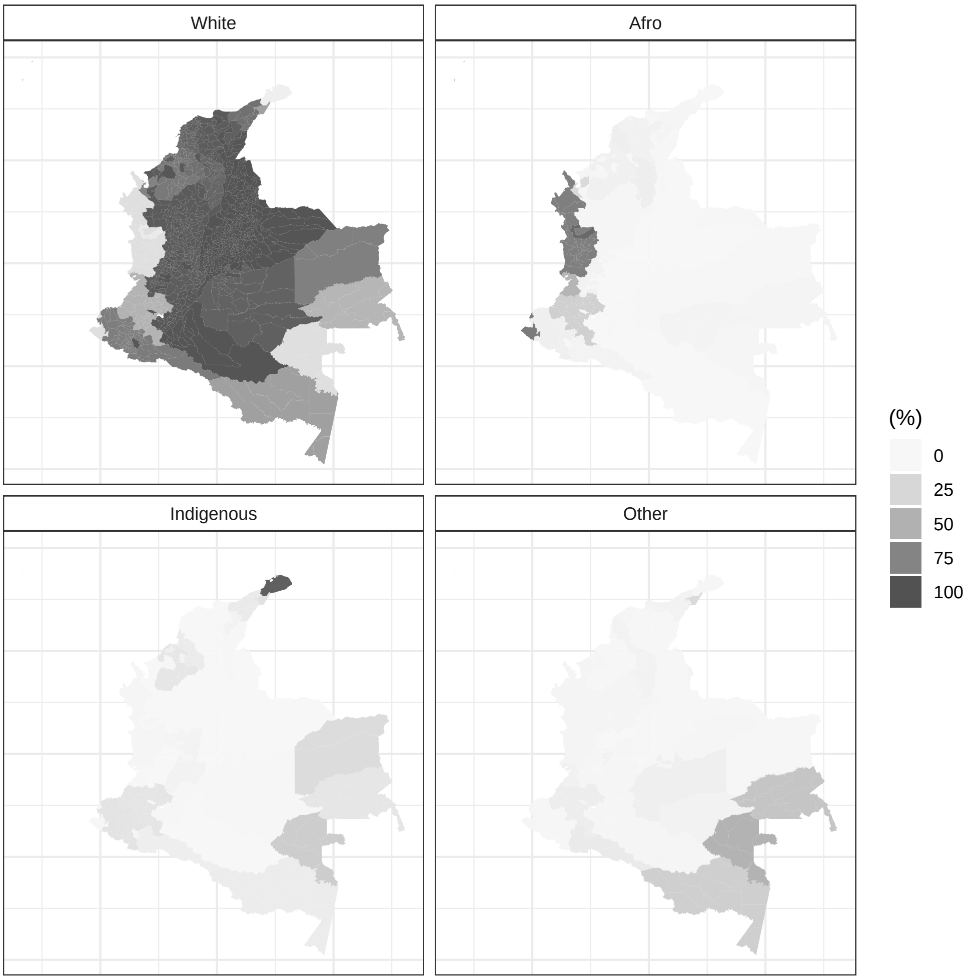

Figure 4.7: Ethnic composition of each LA in 2008

also vary in the way they manage the resources they receive and combine them with their own. Therefore, besides the measures of the availability of resources, chapters 8 and 9 include a fiscal performance indicator to measure how responsibly these resources are managed. The DNP creates this measure using principal component analysis (PCA) with a theoretical scale between 0 and 100, with higher values representing better management practices[14] (DNP, 2008b). In 2008 the index varied between 35 and 84.4.

[14]The fiscal performance index combines the percentage of resources that are spent on running costs, the magnitude of debt, the proportion of spending that is funded by national resources and own resources and the savings ratio.

4.5.2 Conflict and Violence

As described in section 3.2, one of the particularities of Colombia is its long-standing inner conflict. There are two ways in which conflict is officially measured in different regions in Colombia. First, measuring the number of victims of acts of war (such as murder, kidnapping or forced displacement) in a particular area (intensity). Second, measuring the number of victims that arrive from other areas around the country searching for protection (pressure). These two measures are also disaggregated by the perpetrator: guerrillas and organised crime (Red Nacional de Información, 2018).

Chapters 8 and 9 include information on both types of measures of conflict (intensity and pressure) and both types of perpetrator (guerrillas and organised crime). When the perpetrators are guerrillas, the variable name includes 'conflict' (e.g. 'Conflict Int' and 'Conflict Pres' for guerrilla intensity and pressure, respectively), while when the perpetrator is organised crime, the variable name includes 'crime'. In all cases, the indicators are calculated as the number of victims per thousand inhabitants (of their original or recipient region).

There is a large variation in these conflict indicators both over time and across regions. In 2008, there were LAs with no victims of conflict or crime, while on average, there were 13 and 0.3 conflict and crime victims per thousand inhabitants, respectively. Similarly, there were LAs that did not receive victims from the conflict, while on average LAs received 12.6 and 0.2 conflict and crime victims per thousand inhabitants, respectively. Appendix A.2.4 shows the distribution of these variables for all years in the sample.

All the LA variables have been standardised to have a mean of zero and a standard deviation of one in the total sample. This facilitates the interpretation of the models estimated in chapters 8 and 9. Figure 4.8 shows the correlation between the LA variables introduced here, as well as the LA's mean maths test score and its ethnic composition.

As shown in the figure, LAs with a high proportion of White students have a high mean maths score, while the opposite happens in LAs with a high proportion of Afrocolombian, Indigenous and other minority students. LAs with a high average maths score also have better fiscal performance, receive lower transfers to cover running costs and for quality, and are less affected by crime and conflict. The correlation between these LA variables and the proportion of Indigenous students is weak but moderate to

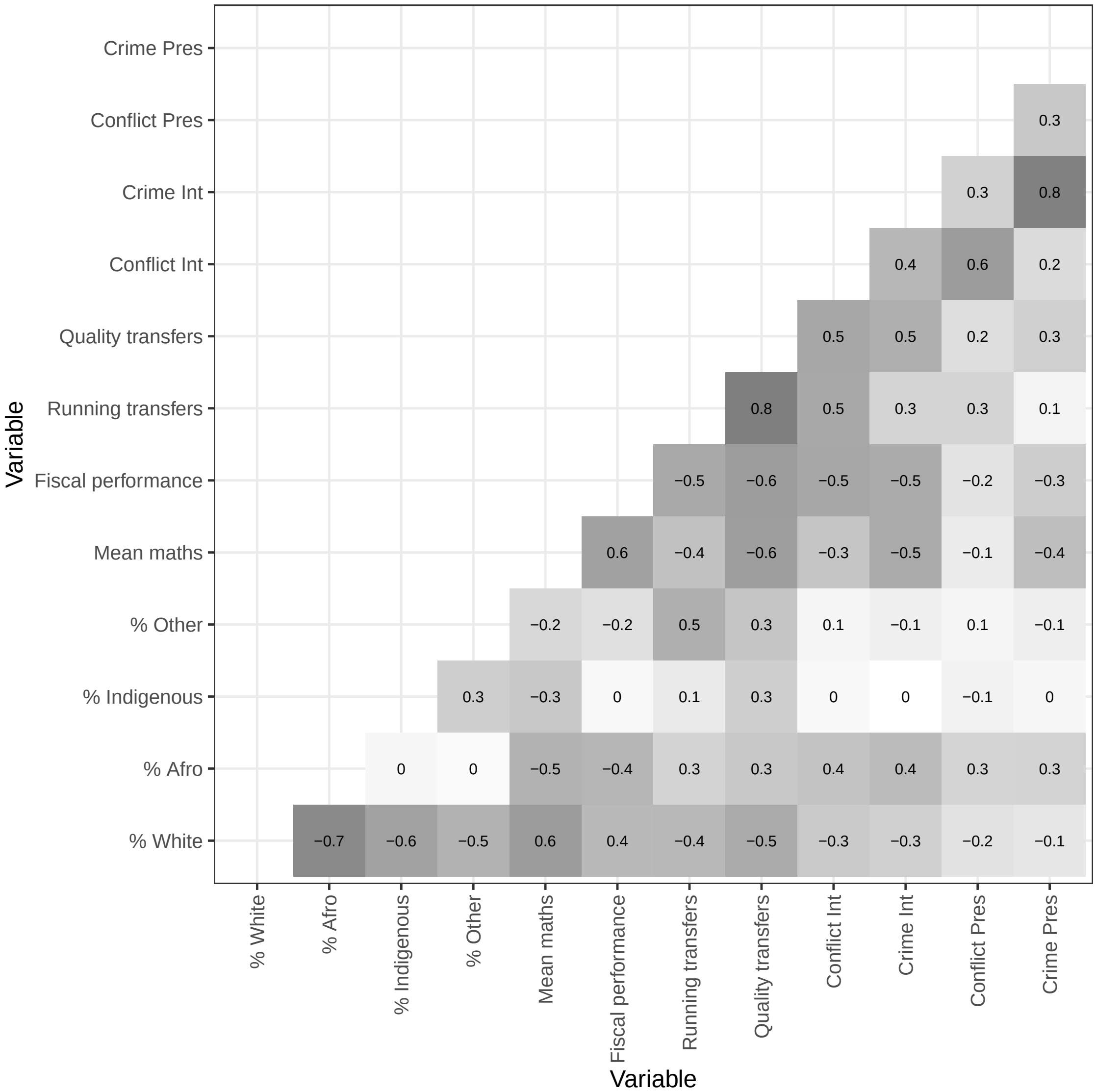

Figure 4.8: Correlation between LA variables in 2008

high for the proportion of White and Afrocolombian students in the LA.

4.6 Data Limitations

The SABER 11 Icfes database includes all students in the country and a rich number of student-level variables. Nonetheless, the database is limited in four ways that restrict the type of analysis that can be conducted. First, students cannot be followed over time, as explained in section 4.2, which deters the study of gaps in learning (achievement growth), value-added by schools, and in general, any studies that require controlling for prior achievement.

Second, data on ethnicity is not supplied for examinations in 3rd, 5th and 9th grades, which means that it is not possible to examine if there is any trend in the ethnic achievement gap as students progress through school, which is the case in the UK and US (Chatterji, 2006; Dustmann et al., 2010; Fryer & Levitt, 2006; Murphy, 2009; Sohn, 2012; Sonnenschein & Galindo, 2014).

Third, the way in which ethnicity is collected does not allow recognising students in the 'other minority' category. They could be unlisted Indigenous groups, immigrants or mixed students. Given that it is the second-largest minority group in the pooled data between 2008 and 2013, with 2.2% of the students in the data (compared to 1.7% for Indigenous students), this implies that this thesis analyses a group of students without knowing who they are.

Finally, even though the database includes around 68 variables about the characteristics of students and their families and some information about schools, there is no information about students' language at home, attitudes towards schooling, parental involvement and expectations, school resources and practices, and about the classrooms. All these variables have been regularly found to be relevant in different studies around the world, as discussed in the literature review chapter (Bodovski & Farkas, 2007; Sonnenschein & Galindo, 2014; Strand, 2011; Vinha, Karino, & Laros, 2016; Wenglinsky, 2004; Wilson et al., 2011). Observing these variables would shed more light on the processes behind ethnic achievement gaps.

4.7 Summary

This chapter presented an overview of the data to gain some intuition behind the results that are presented in the research chapters. It showed that there is an achievement gap between White and minority students, with Afrocolombian, Indigenous and other minority students scoring 0.47 SD, 0.38 SD, and 0.1 SD lower than the majority group in 2008. As further discussed in chapter 6, these gaps are comparable to those reported for minority groups in England (Leckie & Goldstein, 2019; Strand, 2014a) and for Indigenous students in Latin America (Blanco, 2017; McEwan, 2004; McEwan & Trowbridge, 2007), but not as wide as those reported for the US (Page et al., 2008; M. K. Thomas, 2004).

However, the analysis in this chapter also reveals that ethnic minorities come from a more socioeconomically disadvantaged context, larger households and are more likely to be overaged. White and minority students also attend different schools, as White students are more likely to attend private schools, with characteristics as full-day school, academic focus and urban location, that the literature has related to higher achievement. Differences in the school ethnic composition do not only reflect the school's selection processes but also differences in the LA ethnic composition, with White students mainly living in the centre of the country and minority students concentrating in the Caribbean and Pacific coasts of Colombia. These LAs also differ in their availability and management of resources, and the prevalence of conflict. All these differences, at different levels of the education system (students, schools and LAs), are reflected by ethnic achievement gaps. This is why a multilevel approach is required to analysed them, as shown in chapters 6 to 9.

5 Methods

5.1 Introduction

Section 2.7 introduced the conceptual framework for this thesis, understanding the overall ethnic achievement gap as an average that reflects differences between White and minority students with different characteristics and at the different levels of the education system (students, schools and local authorities (LAs)). Chapters 3 and 4 showed that, in Colombia, there are complex relationships between ethnicity and these different levels. This chapter presents an overview of the methods that are employed for separating the effects of these characteristics and levels.

Sections 5.2 and 5.3 introduce the statistical models that allow analysing relationships at different levels and understanding the effects of different characteristics on the gap, respectively. Section 5.4 discusses the limitations of this thesis' methodological approach and section 5.5 summarises. A more thorough review of the methods used in the literature is presented in each research chapter, as appropriate.

5.2 Models for Different Levels

Understanding the ethnic achievement gap as the difference between average test scores implies that a method that allows estimating averages is required, which is why regression analysis is used throughout this thesis. This section describes the single-level and multilevel models that are used with this aim. The explanation focuses on the two-level case of the problem (students within schools), but the intuition can be easily extended to the three-level case (adding LAs). A more general summary of these methods and their assumptions is available in appendix A.3.

5.2.1 Estimating Within- and Between-School Gaps

The overall ethnic achievement gap can be estimated using the simple linear regression model

$$\begin{gathered} y_{ij} = \alpha + \beta M_{ij} + e_{y_{ij}} \\ e_{y_{ij}} \overset{i.i.d.}{\sim} \left(0, \sigma^2\right) \end{gathered} \tag{5.1}$$

where the dependent variable y_{ij} is the student's maths test score and the independent variable M_{ij}, is a dummy variable that equals one for ethnic minority students and nil otherwise. In this case, β represents the average difference in maths test scores between minority and White students: the overall ethnic achievement gap.

Following Burstein (1980), two different sets of regressions can be used to compare students within schools and between schools. The later one can be estimated using the regression model

$$\overline{y}_{.j} = \alpha + \beta^{BS} \overline{M}_{.j} + \overline{e_y}_{.j} \tag{5.2}$$

where the dependent variable is the school mean maths test score $\overline{y}_{.j}$ and the independent variable is the school proportion of minority students $\overline{M}_{.j}$. Thus, β^{BS} represents the mean test score difference between a school with no minority students ($\overline{M}_{.j} = 0$) and a school with no White students ($\overline{M}_{.j} = 1$). β^{BS} does not provide any information about how ethnicity affects individual students, since model (5.2) compares schools, not individuals.

To compare students within the same school, it is necessary to take away the between-school variation from the observations. One way of doing this is by subtracting (5.2) from (5.1). Then,

$$\left(y_{ij} - \overline{y}_{.j}\right) = \alpha + \beta^{W} \left(M_{ij} - \overline{M}_{.j}\right) + \left(e_{y_{ij}} - \overline{e_y}_{.j}\right) \tag{5.3}$$

and thus the dependent $y_{ij} - \overline{y}_{.j}$ and independent variables $M_{ij} - \overline{M}_{.j}$ of this model only vary within schools. Here, β^{W} represents the mean score difference between minority and White students within the same school, i.e. the mean within-school ethnic achievement gap. Therefore, this model shows how the students' ethnicity affects their test scores, but it does not estimate variations in school performance.

An alternative to obtaining within-school achievement gaps is to estimate a model with school fixed-effects, given by

$$y_{ij} = \alpha_j + \beta^W M_{ij} + e_{y_{ij}} \tag{5.4}$$

which includes a dummy variable for every school j (the school fixed-effect α_j) that also takes away the between-school variation. There are two main disadvantages to this model. First, as with (5.3), it is not possible to analyse the effect of any school-level variable on student's achievement, since there would not be any remaining school-level variation which, in the case of model (5.4), translates into a perfect multicollinearity problem with the α_js. Second, model (5.4) can be too computationally demanding, especially when there are many schools in the sample (Grilli & Rampichini, 2011; Shunck, 2010).

Mundlak (1978) demonstrated that, recognising that $M_{ij} = M_{ij} + \overline{M}_{.j} - \overline{M}_{.j}$, it is possible to combine models (5.2) and (5.3) into the model

$$\begin{aligned} y_{ij} &= \alpha + \beta^W \left(M_{ij} - \overline{M}_{.j}\right) + \beta^{BS}\overline{M}_{.j} + e_{y_{ij}} \\ &= \alpha + \beta^W M_{ij} + \left(\beta^{BS} - \beta^W\right)\overline{M}_{.j} + e_{y_{ij}} \\ &= \alpha + \beta^W M_{ij} + \beta^{CS}\overline{M}_{.j} + e_{y_{ij}} \end{aligned} \tag{5.5}$$

where β^W is the within-school achievement gap and $\beta^{CS} = \beta^{BS} - \beta^W$ is the school contextual effect of ethnicity, i.e. the effect of the schools' ethnic composition over and above the students' own ethnicity. Thus, it would be expected that the maths test scores of two (either White or minority) students who attended schools with a ten-percentage-point difference in their proportions of minority students would differ by $\frac{\beta^{CS}}{10}$ standard deviations (SD).

Therefore, model (5.5) is a tool for separating the effects of ethnicity at the student and school level using single-level regression. Not only does it allow separating the within-school gap from the contextual effect of ethnicity, as shown in chapter 6, but also including student and school characteristics to further separate their contribution to the ethnic achievement gap, as shown in chapters 7 and 8, and to study these change for subgroups of students and schools, as shown in chapter 9, by including interaction terms. Adding the LA proportion of minority students allows extending model (5.5) to consider the role of LAs, and multiple ethnic groups can also be considered in the model by including dummy variables for each ethnic minority.

Thus far, it has been assumed that $e_{y_{ij}} \overset{i.i.d.}{\sim} (0, \sigma^2)$, which means that this discussion has been limited to the conventional single-level regression modelling side of the literature. However, this assumption is not realistic in this context since it ignores the correlations between students within the same school. The next section discusses how the multilevel modelling literature has approached this problem and the implications of using this kind of model in the context of this thesis.

5.2.2 Multilevel Modelling

The main difference between single-level and multilevel modelling is that the latter understands the relationships under analysis as generated within clusters. Consequently, when studying the overall (aggregated) relationship, it is necessary to incorporate the uncertainty of each of the relationships within the clusters. In the case of this thesis, academic achievement and the ethnic achievement gap are developed within schools (and LAs when they are also considered) and then estimating an overall achievement gap requires considering the precision with which the mean achievement and the ethnic achievement gaps are estimated for each school.

In practice, this is achieved by including random intercepts and slopes into the linear regression model, which implies that there is a different mean test score and, potentially, a different ethnic achievement gap for each school. Going back to the simple regression model (5.1), in the basic multilevel case, the residual term $e_{y_{ij}}$ is now assumed to be clustered within groups. In that case,

$$\begin{aligned} e_{y_{ij}} &= u_{y_j} + r_{y_{ij}} \\ \\ u_{y_j} &\overset{i.i.d.}{\sim} N\left(0, \sigma_u^2\right) \\ r_{y_{ij}} &\overset{i.i.d.}{\sim} N\left(0, \sigma_r^2\right) \end{aligned} \tag{5.6}$$

leads to a random intercept model that allows each school to have a different mean achievement. In this case, the residuals are correlated with those of students within the same school but uncorrelated with the residuals of students in other schools.

As Rabe-Hesketh and Skrondal (2012) showed, the multilevel model precision-weighted estimate $\widehat{\beta}_{MLM}$ for the overall ethnic achievement gap β can be approximated as the precision-weighted combination of the within- and between-school ethnic achievement gaps $\widehat{\beta}^{W}$ and $\widehat{\beta}^{BS}$

$$\widehat{\beta}_{MLM} \approx \frac{\widehat{\text{Var}}\left(\widehat{\beta}^{W}\right)}{\widehat{\text{Var}}\left(\widehat{\beta}^{W}\right)+\widehat{\text{Var}}\left(\widehat{\beta}^{BS}\right)}\widehat{\beta}^{BS} + \frac{\widehat{\text{Var}}\left(\widehat{\beta}^{BS}\right)}{\widehat{\text{Var}}\left(\widehat{\beta}^{W}\right)+\widehat{\text{Var}}\left(\widehat{\beta}^{BS}\right)}\widehat{\beta}^{W} \tag{5.7}$$

Therefore, if the within-school gap is estimated with higher precision (because there is little variability within schools or school sizes are large), $\widehat{\beta}_{MLM}$ will be closer to the within-school gap; otherwise, $\widehat{\beta}_{MLM}$ will be closer to the between-school gap. This is known as shrinkage (Hox, 2010; Raudenbush & Bryk, 2002). Rabe-Hesketh and Skrondal (2012) noted that the more balanced the data the more accurately (5.7) approximates the multilevel estimate $\widehat{\beta}_{MLM}$. In this thesis, since there is a large variation in the school size (as described in section 4.4), (5.7) is not a good approximation of $\widehat{\beta}_{MLM}$. While some defend that $\widehat{\beta}_{MLM}$ is an accurate representation of the world (e.g. Goldstein, 1995), others have argued that $\widehat{\beta}_{MLM}$ is biased, as it differs, on average, from the unbiased Ordinary Least Squares (OLS) estimate of β (Greene, 2012; Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Chapter 8 explores in more detail the differences in the estimated overall ethnic achievement gap when using single-level and multilevel modelling.

It is therefore normally recommended to use multilevel modelling when analysing parameters at each level (e.g. students and schools) (Begg & Parides, 2003; Bell, Fairbrother, & Jones, 2019; Bell, Jones, & Fairbrother, 2018; Neuhaus & Kalbfleisch, 1998). There are two common approaches to this. First, the use of group-mean centring and second, the inclusion of level-1 aggregated variables (without group-mean centring the level-1 variable), such as the average cluster composition, although Cafri and Fan (2018) showed that using other functions to aggregate the level-1 variables (such as the variance) leads to similar results. Both cases are represented in (5.5). In the case of group-mean centring, the model estimates the within-school gap β^{W} and the between-school gap β^{BS}. In the second case, the model estimates the within-school gap β^{W} and the school contextual effect of ethnicity β^{CS}.

Another point of discussion is the interpretation of β^{CS}. As Castellano, Rabe-Hesketh, and Skrondal (2014) explained, the inclusion of school composition variables is often assumed to represent an underlying process, such as the existence of peer effects (van Ewijk & Sleegers, 2010a), but other times it is just used as a control when the emphasis is in other relationships (Rjosk et al., 2014). Some researchers differentiate between contextual effects, to refer to the characteristics of the context that are directly measured at the cluster level, and composition effects, to refer to level-1 aggregates (e.g. Morris, Manley, & Van Ham, 2018). In this thesis, β^{CS} is interpreted as the school contextual effect of ethnicity, referring to both, the broader school context and its composition. Similar remarks can be made when including the LA ethnic composition into the model.

The models summarised here do not include any other dependent variables. However, these can be easily added to the model, as explored in chapter 8. Multilevel models can also include random slopes (e.g. random ethnic achievement gaps), as presented in chapter 9.

5.3 Effect Decomposition Techniques

In multilevel analysis, it is common to separate the variance of the dependent variable into its within- and between-group components. This variance decomposition is of interest when studying the variation in test scores, for example. A different problem is separating the effect of an independent variable on the dependent variable, for example, when separating the overall ethnic achievement gap into its components, which is the research problem of this thesis. The techniques for approaching this last problem are summarised in this section.

5.3.1 Mediation Analysis

Path and mediation analysis aim to study how variables are linked in a causal path to produce a particular outcome (Agler & De Boeck, 2017), for example, whether students with a higher socio-economic status (SES) have a higher achievement because they have access to a more complete curriculum (Schmidt et al., 2015). In the pursuit of this aim, statistical analysis techniques have been developed to study this type of research problem.

Although it can be argued that the student's ethnicity is not *caused* by any of the variables in the data, the same cannot be said about these other variables or the ethnic composition of schools and LAs. For example, do schools with a higher proportion of minority students score lower on average because of their school composition or because minority students can only access low-quality schools? This kind of question is outside the scope of this thesis, as the available data does not allow differentiating between these two possibilities. The interest is purely in the statistical results developed in the mediation analysis framework, without attaching any causal explanation to the analysis.

To explain the use of mediation analysis in this thesis, please consider the example of exploring[1] how much of the maths achievement gap between White and minority students is attributed to differences in SES. In the model in Figure 5.1, the regression of the student's ethnic group on maths achievement estimates the overall ethnic achievement gap Gap. Then, adding the student's SES to this basic regression model provides an estimate of the ethnic achievement gap, conditional on SES β_{xy}; that is, an ethnic achievement gap that is estimated after comparing White and minority students with the same SES. In this case, the difference between the overall and the conditional ethnic achievement gap $Gap - \beta_{xy}$ represents the part of the overall gap that can be attributed to differences in SES between White and minority students.

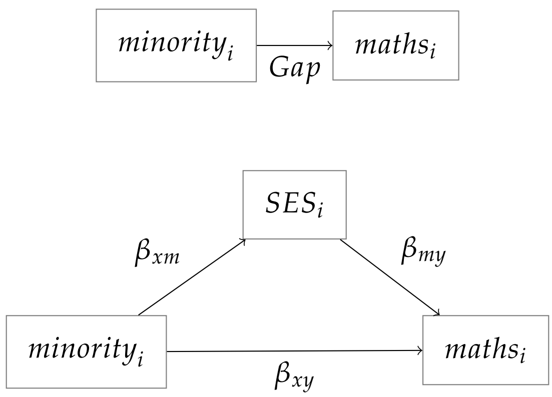

Figure 5.1: Example of mediation model for the ethnic achievement gap

In mediation analysis terms, β_{xy} is the direct effect of ethnicity on maths achievement and $\beta_{xm}\beta_{my} = Gap - \beta_{xy}$ is the indirect of mediated effect.

Therefore, the conditional ethnic achievement gap β_{xy} will be larger than the overall ethnic achievement gap Gap if SES has a positive effect on achievement and minority students tend to have a higher SES, or if SES has a negative effect on achievement and minority students tend to have a lower SES. In turn, the conditional achievement gap β_{xy} will be smaller than the overall ethnic achievement gap Gap if SES has a negative effect on achievement and minority students have a higher SES or if SES has a positive effect on achievement and minority students have a lower SES. A similar line of analysis can be followed to study how other characteristics explain the ethnic achievement gap, as further explored in chapter 8. This result allows exploring the underlying patterns

[1] For a general presentation of mediation analysis, please refer to appendix A.3.3.

behind the overall ethnic achievement gap in chapters 6, 7 and 8.

This basic mediation model has been extended to consider multiple mediators and categorical explanatory variables (Hayes & Preacher, 2014; MacKinnon, 2008). These extensions are further discussed in chapter 6, where they are used to decompose the overall ethnic achievement gap into its components at each level of the education system (students, schools and LAs).

The multilevel version of mediation analysis considers data clustering when studying mediated relationships involving different levels of analysis. For example, whether the schools' socioeconomic composition causes changes in school processes that lead to higher or lower academic achievement for the students (Liu et al., 2015; Rjosk et al., 2014). Authors who developed the methodological framework, however, warn that "the structure of multilevel models dictates that each link in the mediational chain may involve a variable affecting another measured at the same level or a lower level, but not at a higher level"(Krull & MacKinnon, 2001, p. 252-253). Thus, it is not possible to explore if a student-level variable (e.g. ethnicity) affects a school-level variable (e.g. school environment), affecting academic achievement (Preacher, Zyphur, & Zhang, 2010; Tofighi & Thoemmes, 2014). Chapter 6 shows that precisely a model with this structure is a particular case that allows decomposing the ethnic achievement gap into its components at each level of the school system. Another warning in the multilevel extension of mediation models is that the indirect effect of x_i on y_i is not the same when estimated as $\beta_{xm}\beta_{my}$ or as $\beta_t - \beta_{xy}$ due to the precision weights that are implemented in multilevel modelling discussed in section 5.2.2.

5.3.2 The Oaxaca Decomposition

The Oaxaca (1973) decomposition was developed to measure discrimination against women in the US labour market. The intuition behind this technique is that if women with the same education, experience and observable characteristics as men are paid a lower wage, then this difference is due to discrimination. In education, this technique has been used to understand patterns of achievement over time (Barrera-Osorio, García-Moreno, Patrinos, & Porta, 2011; Goussé & Le Donné, 2014) and differences in performance between countries (L. Zhang & Lee, 2011). More importantly, as further discussed in section 3.4, this is the approach that Sánchez-Jabba (2011), the only previous study of the ethnic achievement gap in Colombia, used in the analysis.

In its standard form, the Oaxaca decomposition consists of three steps:

1. Regressing the dependent variable on the set of independent variables for each group, for White and minority students, for example. Assuming that the only relevant variable is SES, this would lead to the models

$$maths_{w_i} = \alpha_w + \beta_w SES_i + e_{w_i} \tag{5.8}$$

$$maths_{m_i} = \alpha_m + \beta_m SES_i + e_{m_i} \tag{5.9}$$

2. Writing the gap as the difference in the predicted values of models (5.8) and (5.9), obtaining

$$Gap = \mathbf{E}\left(maths_{w_i}|w\right) - \mathbf{E}\left(maths_{m_i}|m\right) = \alpha_w + \beta_w \overline{SES}_{wi} - \alpha_m - \beta_m \overline{SES}_{mi} \tag{5.10}$$

3. Since the interest is on the score that minority students would have obtained had the effect of SES been the same for both ethnic groups, transforming (5.10) to include $\beta_w \overline{SES}_{mi}$, which leads to

$$Gap = (\alpha_w - \alpha_m) + \overline{SES}_{mi}\left(\beta_w - \beta_m\right) + \beta_w\left(\overline{SES}_{wi} - \overline{SES}_{mi}\right) \tag{5.11}$$

In (5.11), $\beta_w\left(\overline{SES}_{wi} - \overline{SES}_{mi}\right)$ is called the endowment or explained effect, which is the part of the gap that is attributable to differences in SES between ethnic groups. The remaining part of the gap is unexplained and commonly attributed to discrimination.

The Oaxaca decomposition has been tailored to different applications, including versions for categorical outcomes (Fairlie, 2005), non-linear relationships (Bauer & Sinning, 2008) and a version that uses a pooled regression instead of two separate regressions in step 1 (Juhn, Murphy, & Pierce, 1993). In the achievement gaps literature, Cook and Evans (2000) proposed including school fixed effects (dummy variables for each school) in the estimated models to isolate the within-school gap from differences between schools. This approach has been replicated in different applications around the world, including England (Dustmann et al., 2010), China (Yang et al., 2015), Guatemala (McEwan & Trowbridge, 2007) and other applications in the US (Myers, Kim, & Mandala, 2004; Page et al., 2008). Cook and Evans (2000)'s version has also been extended to include community fixed effects to allow for a more granular decomposition of the gap in Peru (Arteaga & Glewwe, 2019).

Hou (2014) showed that the Oaxaca decomposition is a special case of a mediation analysis model in which the independent variable x_i is binary, and the indirect effect is interpreted as the explained or endowment effect, and the direct effect is interpreted as the unexplained effect or discrimination. Recognising this equivalence allowed Hou (2014) to extend the Oaxaca decomposition to consider gradients (differences in continuous variables) in addition to groupings. This implies that when the overall (unconditional) ethnic achievement gap is compared to the conditional achievement gap (on SES in chapter 7 and on other variables in chapter 6), this analysis is equivalent to performing a Oaxaca decomposition; it allows estimating the part of the gap that is due to differences in the characteristics of White and minority students. In these chapters, however, the method is also extended using contextual-effects models to consider how these characteristics are linked to different levels of the school system (students, schools and LAs).

5.4 Methodological Limitations

The multilevel analysis framework would allow for direct modelling of patterns of change between cohorts attending the same school, which would shed light on the persistence of the ethnic achievement gap within schools. However, the estimation of such a model would require that the test scores vary around the mean from one cohort to the next. This is not the case in the SABER 11 database since the maths test scores had to be standardised for each cohort; the mean maths test score for each cohort is zero, as discussed in chapter 4. For this reason, each model is separately estimated using data for that cohort.

More recent developments in structural equation modelling also allow incorporating corrections for measurement error and missing data (Christ et al., 2017; Marsh et al., 2009; Reardon & Ho, 2015). These are models that add layers of complexity to the models estimated in this thesis. The implications of such models for the study of ethnic achievement gaps are not well known and require further consideration before being incorporated into the framework developed in this thesis. As an alternative, the analysis in each research chapter is subject to robustness checks.

5.5 Summary

This chapter provided an overview of the methods that will be used through the thesis to study the different patterns that are encapsulated by the ethnic achievement gap. Single-level and multilevel regression are coupled with mediation analysis in a pragmatic approach. That is, the models and methods of analysis are viewed as instruments that can be manipulated and used to recognise patterns in the data, according to the conceptual framework for the thesis and the restrictions imposed by the available data. Each research chapter adds a level of complexity to the analysis that allows exploring different patterns of the ethnic achievement gap in Colombia and provides further details on the methods that are employed. Appendix A.3 provides a more general and mathematical presentation of the methods, lists the software that is used in the thesis and the ethical considerations that guide it.

6 How to Decompose the Overall Ethnic Achievement Gaps into Differences Within and Between Schools and Local Authorities? What are the Results of Applying this Technique to the Colombian Context?

6.1 Introduction

As shown in section 3.4, the evidence of ethnic achievement gaps in Colombia is still in its early stage of development, as this problem has only been explored by Sánchez-Jabba (2011). This author estimated a gap of 0.35 standard deviations (SD) in maths test scores between White and minority 11th grade (age 16/17) students who took the SABER 11 exam in the second semester of 2010. Sánchez-Jabba (2011) also found that only 61% of this gap could be explained by the observed variables considered in his study[1]. Therefore, a natural next step is to search for variables that can explain the remaining 39% of the gap. The challenge is to know where to start looking for them, as the international literature reviewed in chapter 2 suggests a wide range of variables at the student, school and higher levels of the school system (Bidwell & Kasarda, 1975; Herman, 2009; Jones et al., 2008; Mohammadpour & Ghafar, 2014; Voight et al., 2015; Wenglinsky, 2009).

One way of tackling this challenge is to understand how the achievement gap can be attributed to different levels of the education system. Once this has been identified, variables at the corresponding levels could be considered as potential explanatory variables of the achievement gap. For example, if most of the achievement gap is due to differences within the schools, the search might start with student-level variables;

[1]These variables are: the region's ethnic composition, the student's household size, whether the student had a job, student's sex, family income, mother's education, school setting (urban or rural), school day type and school fees.

while if most of the achievement gap is a consequence of differences between-schools, school-level variables might be prioritised, including both schools' characteristics and composition. Identifying the contribution of different levels of the education system to the overall achievement gap can also help to prioritise policy efforts at specific levels and to establish what level of the education system may be held accountable for disparities in achievement.

As explained in section 5.3, in social science, there are two dominant ways in which the effect of a variable (e.g. the effect of school composition on achievement) is decomposed. First, in sociology and psychology, the total effect of a variable is decomposed into its direct (e.g., pure peer effects) and indirect effects (e.g the effect of school composition on achievement due to teachers' adjusting their practices) using path and mediation analysis (Alwin & Hauser, 1975; Bodovski & Farkas, 2007; Liu et al., 2015; MacKinnon, Fairchild, & Fritz, 2007). Second, in economics (especially in labour economics), the overall gap between two groups is decomposed into a part that can be explained using observed variables and another part that remains unexplained, using the Oaxaca (1973) decomposition (e.g. Atal, Ñopo, & Winder, 2009; Cook & Evans, 2000; Fryer & Levitt, 2004, 2006; Garbarski, 2015; Martins & Veiga, 2010). Hou (2014) showed that in fact, the Oaxaca (1973) decomposition is equivalent to a mediation analysis decomposition in which the indirect effect is interpreted as the explained part of the gap, and the direct effect is the unexplained part. Based on this finding, Hou (2014) proposed using mediation analysis as a general framework to decompose the effect of an independent variable on a dependent variable.

When the interest is on the achievement gap, an extensive body of research separates the gaps into their within- and between-school parts. However, the focus is rarely on the decomposition itself, but on testing potential explanations for the achievement gap. On one side, researchers try to isolate the effect of the context in an attempt to compare students who are exposed to the same set of environmental influences and thus focus on the within-school achievement gap (Cahan, Barneron, & Kassim, 2014; Cook & Evans, 2000; Fryer & Levitt, 2004, 2006; Rivkin, Hanushek, & Kain, 2005; Stiefel et al., 2007; Voight et al., 2015). On the other side, researchers focus on the effects of group composition, trying to understand its consequences in terms of school effectiveness or to isolate its effects from the influence of self-selection, as is mainly the case for the peer-effects literature and the discussion about ethnic (de)segregation (Angrist, 2014; Lauder, Kounali, Robinson, & Goldstein, 2010; Marsh et al., 2012; Nagengast & Marsh, 2011; Plewis, 2011; Sacerdote, 2011; Schindler Rangvid, 2007; Thrupp, Lauder, & Robinson, 2002; Wilkinson & Fung, 2002).

There is a relatively recent methodological debate about how to interpret estimation results and to attribute differences within- and between-schools to the overall achievement gap. Such debate precedes the recent simultaneous decomposition of the achievement gap into its within- and between-school parts, leading to different conclusions and policy recommendations, ranging from that there is little to be done, to interventions at the school level. In the context of this debate, Hanushek and Rivkin (2006) argued that around 70% of the White-Black achievement gap in the US is due to between-school differences, while Cook and Evans (2000); Fryer and Levitt (2004, 2006); Page et al. (2008); Reardon (2008) argued that these differences only explain at most 40% of the achievement gap. This debate has two limitations: first, it only considers the Black-White achievement gap, ignoring the ethnic achievement gap for other ethnic groups (e.g. Hispanic or Asian) and possible links between ethnic groups. This is a disadvantage as section 4.3 showed that there are differences in the achievement gap by ethnic group, with Afrocolombian students scoring the lowest among all minority groups. Thus, the alternative White-minority comparison may mask the differences among these groups. Second, the debate only focuses on the decomposition of the gap into its within- and between-school components, which ignores other levels of the education system, such as school districts (local authorities (LAs) in Colombia). In the Colombian case, this seems to be a major drawback, as section 4.5 showed that the ethnic composition of LAs varies widely, and this variable is correlated with maths achievement.

This chapter addresses these limitations by taking advantage of the mediation analysis framework as a tool to extend the decomposition of the achievement gap in two ways: first, by considering different ethnic groups instead of only comparing White and minority students; and second, by adding a third level to the decomposition- the LA. By doing this, this chapter provides a more detailed analysis of the ethnic achievement gap that is further explored in the next chapters of this thesis, which search for possible explanations for the achievement gap and its variation among different subgroups.

The remaining of this chapter is organised as follows:

1. Section 6.2 reviews the ethnic gap decomposition debate and the applications that have been developed using the methods in the debate.

2. Section 6.3 shows how the current gap-decomposition debate translates into a mediation analysis framework.

3. Section 6.4 takes advantage of this translation to extend the decomposition of the achievement gap to consider differences at three (or more) levels of the education system and multiple ethnic minority gaps.

4. Section 6.5.1 estimates the overall ethnic achievement gap between White and minority students (pooled together as a single group) and the overall ethnic achievement gaps between White and Afrocolombian, Indigenous and other minority students using the SABER 11 maths test scores between 2008 and 2013.

5. Section 6.5.2 estimates the components of these gaps: the within-school gaps and the school and LA contextual effects of ethnicity.

6. Section 6.5.3 applies the decomposition methodology explained in section 6.4 to decompose the overall ethnic achievement gap into its within-school, between-school/ within LA and between-LA components for 11th grade (age 16/17) students in Colombia.

7. Section 6.6 summarises the results and discusses their limitations and implications for policy and future research

8. Section 6.7 summarises the contributions of this chapter.

6.2 Decomposing the Gap: Debate and Applications

6.2.1 A Recent Methodological Debate

There are three sides to the recent methodological debate about how to decompose the Black-White achievement gap into the parts that can be attributed to differences within and between schools. This debate emerged while attempting to explain the remarkable decrease in the Black-White gap in the US following the abolition of formal segregation. This section introduces the methodological debate to which this chapter contributes. The debate is formalised in section 6.3.

On the first side of the debate are Cook and Evans (2000) and Fryer and Levitt (2004, 2006). Cook and Evans (2000) studied the trends of the US Black-White achievement gap in maths and reading National Assessment of Educational Progress (NAEP) tests for grade 8 and 12 (age 13/14 and 17/18) students. They used a version of the Oaxaca (1973) decomposition that included school fixed effects, allowing them to identify achievement gaps within schools, after considering differences in gender and parental education. Cook and Evans (2000) argued that in 1970 0.3 SD (63%) of the 0.47 SD gap in maths test scores was attributable to differences within-schools, while in 1988 these explained 0.18 SD (59.9%) of the 0.3 SD gap. Based on this finding, Cook and Evans (2000) questioned policy efforts at the school level, including the desegregation policy arguing that "there is little that schools or educational policy can do because families and family background are so important" (Cook & Evans, 2000, p.751).

Fryer and Levitt (2004, 2006) found a gap of 0.52 SD in maths between White and Black students at the end of first grade (age 6/7) in 2000 that broadened to 0.88 SD by the end of third grade (age 8/9) in 2002, using the Early Childhood Longitudinal Study kindergarten cohort (ECLS-K) data. Fryer and Levitt (2004, 2006) also found that when including school fixed-effects in their regressions, there was a 0.02 SD (7.8%) and 0.01 SD (2.4%) reduction on the conditional[2] ethnic achievement gap at the end of first and third grade, respectively. Fryer and Levitt (2004, 2006) focused on the fact that the within-school gap (0.25 SD and 0.37 SD) changed more than the overall ethnic achievement gap (0.27 SD and 0.38 SD) to argue that "the ground lost between first grade and third grade by Blacks is within rather than across schools" (Fryer & Levitt, 2006, p.271).

On the second side of the debate, challenging Fryer and Levitt (2004, 2006), Hanushek and Rivkin (2006) deduced a mathematical formulation to write the Black-White achievement gap as a weighted sum of differences within- and between-schools. Using the ECLS-K, they estimated overall Black-White achievement gaps of 0.15 SD and 0.55 SD in kindergarten and 5th grade (age 10/11 in 2004), respectively. The within-school gaps of 0.13 SD and 0.45 SD in kindergarten and 5th grade accounted for 0.04 SD (25.9%) and 0.07 SD (13.4%) of the overall gap. Based on these findings, Hanushek and Rivkin (2006) concluded that school quality played a major role in the ethic achievement gap and advocated for policies that improved the quality of schools attended by Black students, such as assigning more and better qualified teachers to these schools.

[2]Fryer and Levitt (2004, 2006) controlled for age, birth weight, gender, number of children books at home, mother's age at first birth, socio-economic status (SES), and being a recipient of The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC).

On the third side of the debate, trying to conciliate Fryer and Levitt (2004, 2006) and Hanushek and Rivkin (2006), Reardon (2008) argued that the overall achievement gap can be decomposed into three different components: one that can be attributed to differences within schools, a second one that is due to differences between-schools and a third one that is a combination of both and therefore cannot be uniquely assigned to either of those. Additionally, he argued that the difference between Fryer and Levitt (2004, 2006) and Hanushek and Rivkin (2006) was whether they attributed the third component to the within-school differences (as Fryer and Levitt (2004, 2006)) or to the between-school differences (as Hanushek and Rivkin (2006)). Using the ECLS-K, like previous studies, Reardon (2008) estimated that 0.03 SD (20.9%) of the 0.17 SD overall achievement gap in kindergarten and 0.12 SD (19.9%) of the 0.59 SD overall gap in 5th grade could be attributed to within-school differences. Between-school differences accounted for 0.07 SD (41.1%) and 0.23 SD (39.5%) of the overall gap in kindergarten and 5th grade, respectively and therefore the ambiguous component accounted for 0.07 SD (39.6%) and 0.24 SD (40.6%) of the gap in those grades.

Finally, Page et al. (2008) compared Cook and Evans (2000)'s[3] and Reardon (2008)'s methodology using the reading NAEP test scores for 8th grade (age 13) students. Using Cook and Evans (2000)'s approach, they estimated that 0.87 SD (79%) of the 1.1 SD Black-White achievement gap in 1971 and 0.31 SD (58.8%) of the 0.23 SD gap in 1988 could be attributed to within-school differences, while 0.23 SD (21%) of the gap in 1971 and 0.21 SD (41.2%) in 1988 could be attributed to between-school differences. In turn, using Reardon (2008)'s methodology, Page et al. (2008) estimated that 0.42 SD (38.3%) of the gap in 1971 and 0.15 SD (28.4%) in 1988 could be attributed to the within-school differences, 0.23 SD (20.9%) of the gap in 1971 and 0.21 SD (40.7%) in 1988 to between-school differences and 0.42 SD (38.3%) of the gap in 1971 and 0.15 SD (28.4%) in 1988 could not be unambiguously attributed to either of them. Page et al. (2008) then concluded that the results depended highly on the methodology used for the decomposition. Nonetheless, they favoured Reardon (2008)'s approach because, they argued, eliminating the ambiguous part of the gap would require either a between-school intervention (eliminating segregation) or a within-school intervention (eliminating the within-school gap). Furthermore, Page et al. (2008) combined Reardon (2008)'s with (Oaxaca, 1973) to include

[3]Page et al. (2008) argued that Cook and Evans (2000) and Fryer and Levitt (2004, 2006) followed a similar methodology and therefore focused the debate on Cook and Evans (2000).

covariates in the model. Among the covariates they included are indicator variables for other ethnicities, gender and parental education, finding that exposure to other ethnicities did not affect the Black-White achievement gap. They are the first ones to consider other ethnicities, however, their approach did not consider that the influence of other groups can be part of the between-school component of the Black-White achievement gap, as discussed in section 6.4.2, later in this chapter.

In summary, the three different positions in this debate about the within-between school gap decomposition are: first, Cook and Evans (2000); Fryer and Levitt (2004, 2006) estimated the within-school gap (using fixed school effects) and attributed the rest of the gap to differences between schools. This approach was deemed problematic by all other authors. Second, Hanushek and Rivkin (2006) argued that the gap could be decomposed into its within- and between-school components by considering differences in schools' ethnic composition. Third, Page et al. (2008); Reardon (2008) argued that the gap could not be decomposed only into within- and between-school parts, but that there is a third component that cannot be unambiguously attributed to either of them. Section 6.3.1 explains the different methodologies for the gap decomposition in more detail and argues that each of them leads with a set of parameters with different interpretations.

As further explained in section 5.3, Mundlak (1978) showed that consistent estimates of the within-school achievement gap, that is, of the average test score difference between White and minority students attending the same school, could be achieved by including either school fixed effects (as Cook and Evans (2000); Fryer and Levitt (2004, 2006)) or the schools' mean composition, assuming that ethnicity is not correlated with other (excluded) student-level characteristics. Besides, O. D. Duncan, Cuzzort, and Duncan (1961) (cited by Cronbach, Deken, and Webb (1976)) demonstrated that the overall achievement gap is a weighted sum of both, within- and between-school differences (as Hanushek and Rivkin (2006) argued). Thus, the contribution of each of these components, that is, the part of the overall achievement gap that is explained by within- and between-school differences, is not the same as the within- and between-school achievement gaps themselves.

This chapter puts together these arguments and shows that the three sides of the debate can be captured using a mediation model in which ethnic composition mediates the relationship between ethnicity and achievement. This model is used in this chapter to argue that Hanushek and Rivkin (2006) decomposed the gap into its within-school and between-school components, while other authors present different decompositions of the gap. Furthermore, translating the debate into the mediation analysis framework facilitates extending the decomposition to consider differences at the LA level (and potentially more levels) and the gaps for more ethnic groups, which have received limited attention in the debate so far. The next section summarises some papers that have simultaneously studied within- and between-school differences in achievement.

6.2.2 Related Applications

Outside the ethnic achievement gap literature, the effect (gradient) of SES has been separated into its within- and between-school components, following O. D. Duncan et al. (1961). For example, Causa and Chapuis (2011) used a contextual effects model[4] to differentiate the effect of students' own SES on the Programme for International Student Assessment (PISA) science scores from the contextual effect of this variable for countries in the Organisation for Economic Co-operation and Development (OECD) that participated in PISA 2006. They concluded that, in most countries, the contextual effect was more important than the within-school effect, even after controlling for student, school and country background. With similar findings, Willms (2010) argued in favour of policies to reduce SES segregation between schools to tackle socioeconomic inequalities. However, these studies did not consider ethnic inequalities.

Bohrnstedt, Kitmitto, Ogut, Sherman, and Chan (2015) decomposed the US White-Black achievement gap in maths achievement (measured in NAEP 2011) for 8th grade (age 13/14) students, following Reardon (2008). Bohrnstedt et al. (2015) used a contextual effect model to separate the within-school achievement gap from the contextual effect of ethnicity. They then weighted the within-school achievement gap and the contextual effect to find that while 0.46 SD (51.6%) of the 0.89 SD overall achievement gap was explained by the within-school gap, 0.14 SD (16.1%) by between-school differences, and 0.29 SD (32.3%) could not be explained by either of them.

Quinn (2015b) also followed Reardon (2008)'s approach to decompose the Black-White achievement gap at the end of kindergarten in 2010/2011. However, he performed the decomposition on the conditional ethnic achievement gap , after controlling for a large set of student characteristics, including SES and exposure to the school system. To decompose the Black-White maths achievement gap, he restricted the sample to

[4]Please refer to section 5.2 for a description of this kind of models.

White and minority students, which might result in distorted estimations of the school proportion of Black and White students for each school. Quinn (2015b) found that 0.06 SD (44.4%) of the 0.14 SD conditional Black-White achievement gap could be attributed to differences within schools, 0.005 SD (3.5%) to between-school differences and 0.08 SD (52.1%) to neither of them. Using this result, he argued that there were differences in the school quality attended by White and Black students.

Other studies have adopted Cook and Evans (2000)'s approach of using the Oaxaca (1973) decomposition with school fixed effects (Arteaga & Glewwe, 2019; Dustmann et al., 2010; McEwan, 2004; McEwan & Trowbridge, 2007). Nonetheless, these studies focused on studying how to decompose the gap into a part that could be explained by student characteristics and school (and community in the case of Arteaga and Glewwe (2019)) effects, and a part that remained unexplained. These are discussed in more detail in chapter 8 (section 8.2), which also examines the role of student, school and LA characteristics in explaining the gap.

6.2.3 Group Composition in Mediation Analysis

This chapter shows how mediation analysis can be used as a tool to decompose the overall achievement gap into its within- and between-school components. Such decomposition is achieved by estimating a mediation model in which students' ethnicity is the independent variable, and the school's ethnic composition is the mediator. Thus, this chapter uses mediation analysis as a statistical analysis tool rather than as a method to prove or disprove a causal hypothesis, which is most commonly the case in the literature.

In mediation analysis, authors rarely include an individual's characteristic (like ethnicity) and the group composition according to this characteristic in a mediation model, unless they estimate a multilevel mediation model (Preacher et al., 2010). In this case, researchers estimate a contextual multilevel model, which includes individual characteristics and group composition as independent variables that are mediated by any variable of interest.

For example, Nagengast and Marsh (2012) estimated a model in which academic self-concept mediated the effect of students' and the average classroom achievement on future aspirations. Likewise, Rjosk et al. (2014) estimated a model in which the teachers' focus on language mediated the effect of students' gender, SES, prior achievement, native language and the classroom composition on reading achievement. Similarly, kei

Mak, cheung Cheung, Soh, seong Sit, and kai Ieong (2017) estimated a model to test whether Macao students' reading engagement mediated the within-school gender gap and the contextual effect of gender on reading achievement in PISA 2009. In none of these cases was the group composition used as a mediator. This practice is consistent with the methodological literature in mediation analysis, which asserts that mediator variables can only be measured at the same or lower level than the independent variables, but not at a higher level (Krull & MacKinnon, 2001)

Outside multilevel mediation analysis, when used, group composition usually acts as an independent variable and not as a mediator. This is the case of Liu et al. (2015), who tested whether school processes mediated the relationship between schools' socioeconomic composition and maths achievement in PISA 2003 in OECD countries. An exception is Wodtke and Parbst (2017)'s study that used the school's SES composition as a mediator of the relationship between neighbourhood SES composition and student reading and maths achievement. However, this model still differs from the model discussed in this chapter, as the independent variable is not the students' but the neighbourhood SES. Therefore, the model proposed in this chapter as a device to decompose the ethnic achievement gap has not been estimated in the applied literature either.

6.3 The Debate under the Mediation Analysis Lenses

This section explains how the debate summarised in section 6.2.1, about the achievement gap decomposition into its within- and between-school components, fits into a mediation analysis framework. This debate is limited as it does not consider multiple ethnic groups or higher levels of the school system (such as LAs). Thus, building on section 5.2.1 in the methods chapter, which describes the linear regression models through which estimates of the within- and between-school achievement gaps can be obtained, this section first briefly summarises how authors on the three sides of the debate decomposed the achievement gap and then shows how mediation analysis encompasses these decomposition strategies.

6.3.1 Decomposing the Overall Gap into its Within-School and Between-School Components

Section 6.2.1 showed that there are three ways in which the overall achievement gap β has been decomposed into its within- and between-school components:

1. Cook and Evans (2000) and Fryer and Levitt (2004, 2006) estimated the fixed-effects model $y_{ij} = \alpha_j + \beta^W M_{ij} + e_{y_{ij}}$ (5.4), where y_{ij} is the student's maths test score, M_{ij} is a dummy variable that equals one for ethnic minority students and zero otherwise, α_j is a dummy variable for every school j and $e_{y_{ij}} \overset{i.i.d.}{\sim} \left(0, \sigma^2\right)$ is a student-level error term. In this case, β^W is the within-school ethnic achievement gap; i.e. the difference in mean test scores between White and minority students attending the same school. Cook and Evans (2000) and Fryer and Levitt (2004, 2006) attributed β^W to within-school differences and the remainder of the gap ($\beta - \beta^W$) to between-school differences.

2. Hanushek and Rivkin (2006) represented the decomposition using a formula that is equivalent to using the parameters estimated in the between-school model $\overline{y}_{.j} = \alpha + \beta^{BS}\overline{M}_{.j} + \overline{e_y}_{.j}$ (5.2) and the within-school model $\left(y_{ij} - \overline{y}_{.j}\right) = \alpha + \beta^W \left(M_{ij} - \overline{M}_{.j}\right) + \left(e_{y_{ij}} - \overline{e_y}_{.j}\right)$ (5.3) to calculate

$$\beta = (1-\omega)\,\beta^W + \omega\beta^{BS}, \omega = \frac{\mathbf{Var}\left(\overline{M}_{.j}\right)}{\mathbf{Var}\left(M_{ij}\right)} \tag{6.1}$$

   where $\overline{M}_{.j}$ is the school proportion of minority students and $\overline{y}_{.j}$ is the school mean test score, therefore β^{BS} represents the mean achievement difference between a school with no minority students and a school with no White students. Then, Hanushek and Rivkin (2006) argued that the contribution of the within-school achievement gap β^W to the overall gap β is $(1-\omega)\,\beta^W$ and the contribution of the between-school differences β^{BS} to the overall achievement gap is $\omega\beta^{BS}$, as had already been shown by Burstein (1980); Cronbach et al. (1976).

3. Reardon (2008) and Page et al. (2008) estimated the contextual-effect model $y_{ij} = \alpha + \beta^W M_{ij} + \beta^{CS}\overline{M}_{.j} + e_{y_{ij}}$ (5.5), where $\beta^{CS} = \beta^{BS} - \beta^W$ is the school contextual effect of ethnicity, i.e. the effect of the schools' ethnic composition over and above the students' own ethnicity. Reardon (2008) and Page et al. (2008) decomposed the gap in three parts: the within-school part $(1-\omega)\,\beta^W$, the between-school part

$\omega\beta^{CS}$ and the ambiguous part $\omega\beta^{W}$. Reardon (2008) demonstrated that he and Hanushek and Rivkin (2006) used the same weight ω to decompose the gap and that this weight is the difference in the mean school proportion of minority students for minority and for White students, that is:

$$\omega = \mathrm{E}\left[\overline{M}_{.j}|M_{ij} = 1\right] - \mathrm{E}\left[\overline{M}_{.j}|M_{ij} = 0\right] \tag{6.2}$$

Furthermore, Reardon (2008, p.23) noted that ω is equivalent to the variance ratio index of segregation. A value of zero means that there is no segregation; i.e. White and minority students attend schools with the same ethnic composition, and a value of one means complete segregation[5]. Table 6.1 summarises this debate by showing the components of the overall achievement gap β, according to each of the authors.

Table 6.1: Summary of the achievement gap decomposition debate

Authors	Model	Component		
		Within-School	Between-School	Ambiguous
Cook and Evans (2000); Fryer and Levitt (2004, 2006)	(5.4)	β^{W}	$\beta - \beta^{W}$	
Hanushek and Rivkin (2006)	(6.1)	$(1-\omega)\,\beta^{W}$	$\omega\beta^{BS}$	
Page et al. (2008); Reardon (2008)	(5.5)	$(1-\omega)\,\beta^{W}$	$\omega\beta^{CS}$	$\omega\beta^{W}$

$\omega = \mathrm{E}\left[\overline{M}_{.j}|M_{ij} = 1\right] - \mathrm{E}\left[\overline{M}_{.j}|M_{ij} = 0\right]$

To interpret the differences between these decompositions, Reardon (2008) and Page et al. (2008) proposed analysing the type of policies that would be required to eliminate ethnic differences within and between schools. If eliminating a component of the gap requires a between-school intervention, it can be attributed to the between-school gap. Conversely, if it requires a within-school intervention, it can be attributed to the within-school component of the gap. The argument is based on three different policies:

[5]Segregation is normally understood to include five dimensions: evenness, concentration, centralisation, clustering and exposure (Massey & Denton, 1988). ω is a measure of this last dimension.

1. Eliminating the within-school gap ($\beta^W = 0$) while leaving segregation (ω) and the school mean achievement unchanged[6]: This within-school effort would lead to a decrease of $(1-\omega)\,\beta^W$ in the overall gap β, which now is $\beta = \omega\beta^{CS}$, composed only of differences between schools. Therefore, Reardon (2008) and Page et al. (2008) argued that $(1-\omega)\,\beta^W$ is the part of the gap that can 'unambiguously' be attributed to within-school differences, in agreement with Hanushek and Rivkin (2006).

2. Eliminating segregation ($\omega = 0$): This between-school effort would lead to a decrease of $\omega\beta^{CS}$ in the overall gap, which would now only be the within-school gap $\beta = \beta^W$, as argued by Fryer and Levitt (2004, 2006). Therefore, Reardon (2008) and Page et al. (2008) interpreted $\omega\beta^{CS}$ as the part of the gap that is attributable to segregation and 'unambiguously' to between-school differences.

3. Eliminating the relationship between the school ethnic composition and its mean achievement $\beta^W + \beta^{CS}$ while keeping the within-school gap β^W and segregation ω at the same levels[7]: This between-school effort would lead to a reduction of $\omega\left(\beta^W + \beta^{CS}\right)$ in the overall gap β, which would now be $\beta = (1-\omega)\,\beta^W$; the 'unambiguously' within-school component of the gap, which Reardon (2008) and Page et al. (2008) argued is in agreement with Hanushek and Rivkin (2006)'s decomposition.

Using these policy scenarios, Reardon (2008) and Page et al. (2008) argued that $\omega\beta^W$ cannot be 'unambiguously attributed' to either the within- or the between-school component of the gap because it is an interaction between within-school differences β^W and segregation ω, which can be changed through interventions either within or between schools. The difference then, they argued, between Fryer and Levitt (2004, 2006)'s and Hanushek and Rivkin (2006)'s decompositions is that the first ones attributed $\omega\beta^W$ to the within-school component, while the second ones attributed it to the between-school component. This argument can be questioned[8] by noticing that the 'unambiguously

[6]Please note that in this scenario, leaving the school mean unchanged requires an increase of β^W in the school contextual effect of ethnicity β^{CS}.

[7]Please note that such policy is only possible if $\beta^W = -\beta^{CS}$; i.e. if the within-school gap offsets the school contextual effect of ethnicity.

[8]This proposition does not question the decomposition itself, but the argument used to decide how to interpret each of the components.

within-school' component of the gap $(1-\omega)\,\beta^W$ could also be eliminated through a between-school intervention: completely segregating schools ($\omega = 1$). Unlike Reardon (2008) and Page et al. (2008), this chapter argues that these are different decompositions of the overall ethnic achievement gap, which can be more easily understood when translating the debate into a mediation analysis framework, which is the task of next section.

6.3.2 Translating the Debate into Mediation Analysis

After Reardon (2008)'s presentation of the decomposition of the achievement gap into its within- and between-school and ambiguous components, it is relatively easy to understand how this can be translated into a mediation analysis framework. The direct/indirect effect decomposition can then be used as a device to extend the current decomposition to consider additional groups and levels of the school system. It is enough to consider a model in which the ethnic composition of each school $\overline{M}_{.j}$ mediates the relationship between ethnicity M_{ij} and maths achievement y_{ij}, as the one shown in Figure 6.1.

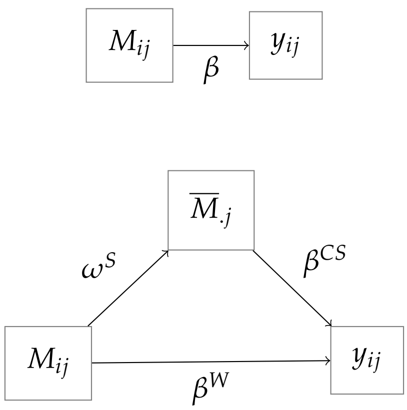

Figure 6.1: Model for the achievement gap decomposition into its within- and between-school components

The upper part of Figure 6.1 describes the total effects model $y_{ij} = \alpha + \beta M_{ij} + e_{y_{ij}}$ (5.1), where β represents the overall ethnic achievement gap between minority and White students. The lower part of Figure 6.1 represents the outcome and mediation model. The outcome model is the contextual effect model $y_{ij} = \alpha + \beta^W M_{ij} + \beta^{CS}\overline{M}_{.j} + e_{y_{ij}}$ (5.5) in which β^W is the within-school achievement gap and β^{CS} is the contextual effect of ethnicity at the school level. Besides, the mediation model is

$$\overline{M}_{.j} = \gamma + \omega^S M_{ij} + e_{S.j}, \forall i \in j \tag{6.3}$$

where the school proportion of minority students $\overline{M}_{.j}$ is repeated for all students i in school j. Here, (6.3) is a simple linear regression model that is used as a device to compute the segregation index ω in (6.2), as ω^S is the difference in the mean school ethnic composition between minority and White students, given by (6.2). If $\omega^S = 0$, White and minority students attend schools with the same proportion of minority students, while if $\omega^S = 1$ White students only attend schools with White students and minority students only attend schools with minority students.

In the mediation analysis context, the total effect β is decomposed into a direct effect β^W and an indirect effect $\omega^S \beta^{CS}$, such that

$$\beta = \beta^W + \omega^S \beta^{CS} \tag{6.4}$$

Here, the direct effect is the within-school gap β^W, that is, the average difference in test scores between White and minority students attending the same school, and the indirect effect $\omega^S \beta^{CS}$ represents how much more minority students are affected by the school contextual effect of ethnicity β^{CS}, in comparison to White students, given the current level of between-school segregation ω^S. Cook and Evans (2000) and Fryer and Levitt (2004, 2006) attributed β^W to the within-school component of the gap and $\omega^S \beta^{CS}$ to the between-school component of the gap.

The expression in (6.4) shows that Cook and Evans (2000)'s and Fryer and Levitt (2004, 2006)'s interpretation implies that eliminating the within-school gap $\beta^W = 0$ would leave the contribution of the between-school differences to the gap unaffected. This statement is true in a scenario without segregation ($\omega^S = 0$), where the ethnic achievement gap β is fully explained by the within-school achievement gap ($\beta = \beta^W$). In this case, there are no differences in the average test scores of schools typically attended by White and minority students. However, in a scenario with segregation ($\omega^S \neq 0$), suppressing the within-school ethnic achievement gap ($\beta^W = 0$) leads to changes in ethnic differences both within and between schools, which the between-school component of this decomposition does not capture. Therefore, Cook and Evans (2000) and Fryer and Levitt (2004, 2006) decomposed the achievement gap into its direct and indirect effects; the within-school gap β^W and the effect of segregation $\omega^S \beta^{CS}$, rather than into its within- and between-school components.

Translating the mediation analysis decomposition into a within- and between-school gap decomposition requires recognising that differences between schools are the sum of differences within schools and the contextual effect of ethnicity. In other words, differences between schools arise not only because there is an effect of studying with a larger proportion of minority students (β^{CS}, a school-level effect) but also because of differences between students of different ethnicity across all schools (β^{W}). Equivalently, this can be expressed as $\beta^{CS} = \beta^{BS} - \beta^{W}$. Therefore,

$$\beta = \left(1 - \omega^{S}\right) \beta^{W} + \omega^{S} \beta^{BS} \tag{6.5}$$

which is equivalent to Hanushek and Rivkin (2006) decomposition, with $\left(1 - \omega^{S}\right) \beta^{W}$ corresponding to the within-school component of the gap and $\omega^{S} \beta^{BS}$ to the between-school component of the gap. According to (6.5), the ethnic achievement gap β is a weighted sum of differences within schools β^{W} and between schools β^{BS}. In the scenario without segregation ($\omega^{S} = 0$), this decomposition is equivalent to Cook and Evans (2000)'s and Fryer and Levitt (2004, 2006)'s, with $\beta = \beta^{W}$. If there is segregation ($\omega \neq 0$), supressing the within school ethnic achievement gap ($\beta^{W} = 0$) implies changes in both, the within-school $\left(1 - \omega^{S}\right) \beta^{W}$ and the between-school $\omega^{S} \beta^{BS}$ components of the gap.

To further separate the changes in the between-school component of the gap, it is enough to recognise that $\beta^{BS} = \beta^{W} + \beta^{CS}$ in (6.5), which leads to

$$\beta = \left(1 - \omega^{S}\right) \beta^{W} + \omega^{S} \beta^{W} + \omega^{S} \beta^{CS} \tag{6.6}$$

which is equivalent to Reardon (2008)'s decomposition, where, under his interpretation, $\left(1 - \omega^{S}\right) \beta^{W}$ is the within-school component of the gap, $\omega^{S} \beta^{CS}$ is the between-school component, due to segregation, and $\omega^{S} \beta^{W}$ is the part of the gap that cannot be unambiguously attributed to either of them. Reardon (2008) and Page et al. (2008) argued that $\omega^{S} \beta^{W}$ is an ambiguous part of the gap because eliminating it requires a combination of both, within- and between-school interventions. In turn, the argument in this thesis is that $\omega^{S} \beta^{W}$ is part of the between-school component of the gap, as argued by Hanushek and Rivkin (2006), since differences between schools involve both a part that reflects differences between students across all schools (β^{W}) and a part that reflects the effect of the ethnic composition of each school (β^{CS}). In that sense, Reardon (2008)'s and Page et al. (2008)'s is not a within- and between-school achievement gap decomposition, but a decomposition in three parts: the contribution of within-school differences $\left(1-\omega^S\right)\beta^W$ at the student level, the contribution of segregation through differences in the student intake $\omega^S\beta^W$, and the contribution of segregation through the school contextual effect of ethnicity $\omega^S\beta^{CS}$.

As the objective of this chapter is to decompose the overall gap into its within- and between-school and LA components, Reardon (2008)'s and Page et al. (2008)'s disaggregation of the between-school component of the gap is unnecessary here and therefore the overall ethnic achievement gap β will be decomposed as in (6.5). However, this methodology needs to be extended to consider the LA level before it is possible to apply it to achieve this objective.

6.4 Extending the Gap Decomposition using Mediation Analysis

Translating the ethnic gap decomposition debate into a mediation analysis framework unlocks possible extensions of the basic decomposition to consider more complex mediation models. This section presents two extensions of the achievement gap decomposition based on this framework. The first one includes an extra level of analysis (LA), and the second one allows comparisons between more than two groups (i.e., analysing the different ethnic groups instead of comparing all minority students). Both extensions are based on Hayes and Preacher (2014) description of mediation analysis with multiple mediators and a multicategorical independent variable.

6.4.1 Adding Extra Decomposition Levels

Extending section 6.3, the achievement gap can be decomposed into its within-, between-school and between-LA components using a model in which the school and LA proportions of minority students, $\overline{M}_{.jk}$ and $\overline{M}_{..k}$ mediate the relationship between the student's ethnicity M_{ijk} and maths test scores y_{ijk}. The total effects model in this case is $y_{ijk}=\alpha+\beta M_{ijk}+e_{y_{ijk}}$ with $e_{ijk}\sim$ i.i.d. $\left(0,\sigma^2\right)$, which is similar to (5.1) as β represents the overall achievement gap between minority and White students. Figure 6.2 represents the outcome and the mediation models.

The outcome model is the contextual effect model

$$y_{ijk}=\alpha+\beta^W M_{ijk}+\beta^{CS}\overline{M}_{.jk}+\beta^{CL}\overline{M}_{..k}+e_{y_{ijk}} \tag{6.7}$$

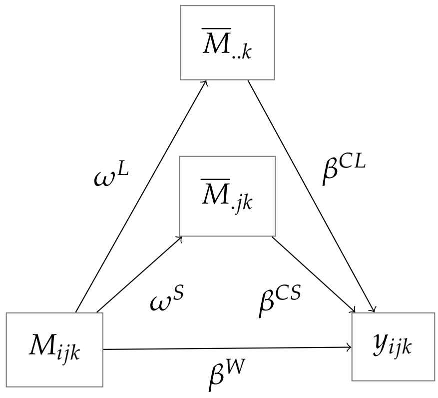

Figure 6.2: Model for the achievement gap decomposition into its within-, between-school and between-LA components

where β^W is the within-school achievement gap, β^{CS} is the school contextual effect of ethnicity and β^{CL} is the LA contextual effect of ethnicity. Thus, students attending schools with a ten percentage points higher proportion of minority students are expected to score $\frac{\beta^{CS}}{10}$ more, independently of their own ethnicity and LA's ethnic composition. Similarly, students in LAs with a ten percentage points higher proportion of minority students are expected to score $\frac{\beta^{CL}}{10}$ more, independently of their ethnicity. This implies that β^{CS} is now the school contextual effect of ethnicity within LAs.

Since the model includes two mediators, there are two mediation equations

$$\begin{aligned} \overline{M}_{.jk} &= \gamma_1 + \omega^S M_{ijk} + e_{S.jk}, \forall i \in j,k \\ \overline{M}_{..k} &= \gamma_3 + \omega^L M_{ijk} + e_{L..k}, \forall i \in k \end{aligned} \tag{6.8}$$

Again, (6.7) and (6.8) are used as devices to decompose the ethnic achievement gap rather than as models with a particular interpretation. In particular, (6.8) allows estimating the between-school segregation index ω^S and the between-LA segregation index ω^L. Notice that considering the LA level does not imply that the between-school segregation index ω^S is conditional to the LA, as it is the same as in (6.3).

Following Hayes and Preacher (2014), given these relationships, the overall gap β can be decomposed as

$$\beta = \beta^W + \omega^S \beta^{CS} + \omega^L \beta^{CL} \tag{6.9}$$

where, again, β^W would be interpreted as the direct effect and $\omega^S \beta^{CS}$ and $\omega^L \beta^{CL}$ as the parts of the total indirect effect corresponding to the school and LA ethnic composition, respectively. However, $\beta^{CS} = \beta^{BS} - \beta^W$ and $\beta^{CL} = \beta^{BL} - \beta^{BS}$, that is, the school contextual effect is the difference in test scores related to a larger proportion of minority students net of the effect of the students' own ethnicity, and the LA contextual effect of ethnicity is equivalent to the differences in mean test scores between LAs with different ethnic composition net of the effect of differences between schools . Recognising this in (6.9) allows writing the gap as a combination of within-school, between-school/within-LA and between-LA components as

$$\beta = \left(1 - \omega^S\right) \beta^W + \left(\omega^S - \omega^L\right) \beta^{BS} + \omega^L \beta^{BL} \tag{6.10}$$

where $\left(1 - \omega^S\right) \beta^W$ is the within-school component of the gap, $\left(\omega^S - \omega^L\right) \beta^{BS}$ is the between-school/within LA component and $\omega^L \beta^{BL}$ is the between LA component of the gap.

This decomposition and the one presented in section 6.3.2 assume that the gap between Afrocolombian and White students is the same as the gap between Indigenous and White students and that how the gap can be attributed to differences between students, schools and LAs is the same for all ethnic groups. This assumption is probably incorrect, given the differences in test scores among ethnic groups presented in the data analysis chapter (section 4.3). Therefore an alternative decomposition that allows for different gaps and components for each ethnic group may be more attractive. This decomposition is explored in the next section.

6.4.2 Comparing More Groups

The second extension to the achievement gap decomposition in this chapter allows for the inclusion of categorical variables with more than two categories. In this case, there are three possible ethnic categories besides White, which is the reference category: Afrocolombian A_{ij}, Indigenous I_{ij}, and other O_{ij}. This matches the definition of the ethnic achievement gap in this thesis (see section 2.2), which compares the achievement of minority groups with that of White students. Should the comparison of interest be between any other two groups (e.g. Afrocolombian versus Indigenous students), it is enough to change the reference category (e.g. to Afrocolombian students).

This extension can be combined with the extension discussed in section 6.4.1 to decompose the achievement gaps of the different ethnic groups into their within-, between-school and LA components, as explained in appendix A.4.1. However, to facilitate the explanation, this section limits the presentation to the within- and between-school decomposition.

Figure 6.3 depicts the total effect model when different ethnic groups are considered. Here, each ethnic group is represented by a dummy variable that equals one when a student belongs to that ethnic group and nil otherwise. All dummy variables are independent variables, and each one is associated with a unique gap between students of that ethnic group and White students. Therefore,

$$y_{ij} = \alpha + \beta_A A_{ij} + \beta_I I_{ij} + \beta_O O_{ij} + e_{y_{ij}} \tag{6.11}$$

where β_A, β_I and β_O represent the gasp in maths test scores between White and Afrocolombian, Indigenous, other ethnic minorities, respectively.

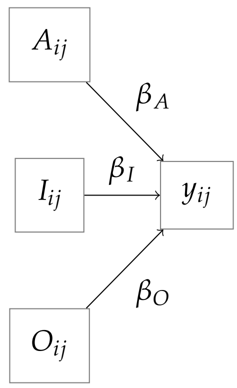

Figure 6.3: Total effect model for the decomposition of the gaps for different ethnic groups

The outcome and mediation models are represented in Figure 6.4, which follows Hayes and Preacher (2014) explanation. This shows that besides having an explanatory variable for each minority group, there is also one mediator for each of them. These mediators are the school proportion of students belonging to each of the minority groups[9], $\overline{A}_{.j}$, $\overline{I}_{.j}$ and $\overline{O}_{.j}$. Consequently, the outcome model corresponds to the contextual effect model

$$y_{ij} = \alpha + \beta_A^W A_{ij} + \beta_I^W I_{ij} + \beta_O^W O_{ij} + \beta^{CSA}\overline{A}_{.j} + \beta^{CSI}\overline{I}_{.j} + \beta^{CSO}\overline{O}_{.j} + e_{y_{ij}} \tag{6.12}$$

[9]Should the comparison of interest be with respect to other ethnic group (e.g. Afrocolombians), the mediators would be the school proportions of students in the remaining categories (e.g. White, Indigenous and other minorities).

where β_A^W, β_I^W and β_O^W are the within-school achievement gaps for each group and β^{CSA}, β^{CSI} and β^{CSO} are the school contextual effects of each minority group, conditional on the school proportion of other ethnic minorities.

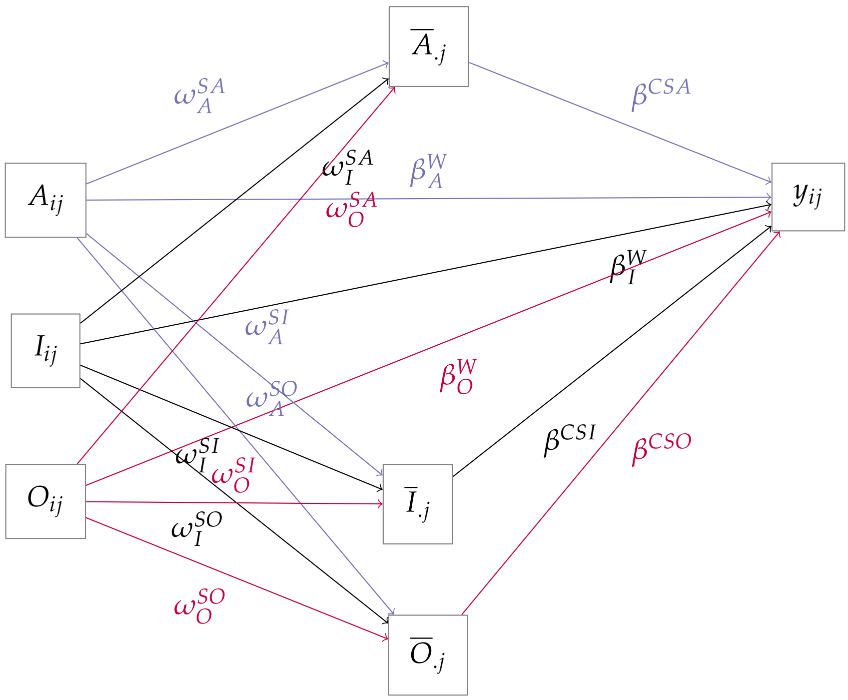

Figure 6.4: Outcome and mediation models for the decomposition of the gaps for different ethnic groups

Following Hayes and Preacher (2014), there is one mediation model for each mediator, and thus for each ethnic minority group, and each mediator is a function of the school ethnic composition in terms of all minority groups. As in sections 6.3.2 and 6.4.1, this is only a device to decompose the achievement gap for each minority group, which results in the models

$$
\begin{aligned}
\overline{A}_{.j} &= \gamma_1 + \omega_A^{SA} A_{ij} + \omega_I^{SA} I_{ij} + \omega_O^{SA} O_{ij} + e_{SA.j}, \forall i \in j \\
\overline{I}_{.j} &= \gamma_2 + \omega_A^{SI} A_{ij} + \omega_I^{SI} I_{ij} + \omega_O^{SI} O_{ij} + e_{SI.j}, \forall i \in j \\
\overline{O}_{.j} &= \gamma_3 + \omega_A^{SO} A_{ij} + \omega_I^{SO} I_{ij} + \omega_O^{SO} O_{ij} + e_{SO.j}, \forall i \in j
\end{aligned}
\tag{6.13}
$$

Again the mediator models provide estimates of indicators of segregation, given by the differences in the mean school proportion of students of each minority group for each minority and White students. For instance, ω_A^{SI} is the difference in the mean school proportion of Indigenous students between Afrocolombian and White students; if $\omega_A^{SI} = 0$, White and Afrocolombian students attend schools with the same proportion of Indigenous students.

Therefore, following Hayes and Preacher (2014), each of the gaps can be decomposed into their direct and indirect effects analogously to (6.9) in the previous section. There is a direct effect for each ethnic minority, which corresponds to the within-school achievement gap. Each of the mediators contributes to the total indirect effects for each ethnic group. Importantly, this implies that the achievement gap of each ethnic group is not only a function of its own contextual effect but also a function of the contextual effects of other minority groups. For example, the overall ethnic achievement gap for Afrocolombians can be decomposed into the within-school gap for Afrocolombians β_A^W (the direct effect) and the effect of school context $\omega_A^{SA}\beta^{CSA} + \omega_I^{SA}\beta^{CSI} + \omega_O^{SA}\beta^{CSO}$ (the indirect effect). In general,

$$\begin{aligned}
\beta_A &= \beta_A^W + \omega_A^{SA}\beta^{CSA} + \omega_I^{SA}\beta^{CSI} + \omega_O^{SA}\beta^{CSO} \\
\beta_I &= \beta_I^W + \omega_A^{SI}\beta^{CSA} + \omega_I^{SI}\beta^{CSI} + \omega_O^{SI}\beta^{CSO} \\
\beta_O &= \beta_O^W + \omega_A^{SO}\beta^{CSA} + \omega_I^{SO}\beta^{CSI} + \omega_O^{SO}\beta^{CSO}
\end{aligned} \tag{6.14}$$

Recognising that the between-school gap for each ethnicity depends on the within-school gap and contextual effect of such ethnicity (e.g., $\beta^{BSA} = \beta_A^W + \beta^{CSA}$) into the mediation analysis decomposition, the overall achievement gap for each minority group can be decomposed as

$$\begin{aligned}
\beta_A &= \left(1 - \omega_A^{SA}\right)\beta_A^W + \omega_A^{SA}\beta^{BSA} + \omega_A^{SI}\beta^{CSI} + \omega_A^{SO}\beta^{CSO} \\
\beta_I &= \left(1 - \omega_I^{SI}\right)\beta_I^W + \omega_I^{SI}\beta^{BSI} + \omega_I^{SA}\beta^{CSA} + \omega_I^{SO}\beta^{CSO} \\
\beta_O &= \left(1 - \omega_O^{SO}\right)\beta_O^W + \omega_O^{SO}\beta^{BSO} + \omega_O^{SA}\beta^{CSA} + \omega_O^{SI}\beta^{CSI}
\end{aligned} \tag{6.15}$$

Therefore, the overall achievement gap for each ethnicity cannot be decomposed into its within-school and between-school components without taking into account the contextual effects of the other minority groups, which are weighted according to their relative (to that of White students) exposure ω to other ethnic groups. For example, the overall achievement gap for Afrocolombian students does not only depend on the within-school and between-school Afrocolombian achievement gap but also the influence of the context by Indigenous and other minority students. The empirical attempts to incorporate several groups into the analysis (discussed in section 6.2.2) have not recognised the role of other ethnic groups as part of a school-level element of the gap (Dustmann et al., 2010; Page et al., 2008; Quinn, 2015b). The importance of this omission depends on how segregated minority groups are with respect to each other and how strong their contextual effects are. Alternatively, performing pair-wise comparisons between white and each minority group (e.g. Afrocolombian) potentially brings additional distortions into the analysis. Performing the analysis by restricting the sample to include these two groups (e.g. White and Afrocolombian) of students incorporates distortions in the estimated school proportion of minority students, by reducing the total number of students within each school, and thus into the decomposition analysis. The alternative of focusing only on schools that are exclusively attended by each pair of ethnic groups (e.g. White and Afrocolombian) might exclude a large proportion of school and therefore ignore a substantial source of variation when estimating between-school gaps.

6.5 Results

The previous sections described how to extend the current methodology to decompose the overall ethnic achievement gap into its within-school, between-school/within-LA and between-LA components. This section applies these decompositions to the SABER 11 data described on chapter 4. The cohorts between 2008 and 2013 are examined separately, with each set of models being estimated for each cohort.

Page et al. (2008) suggested excluding schools in which all students belong to the same ethnic group (i.e. ethnically homogeneous schools) because these do not contribute to the estimation of the within-school gap. However, in Colombia, this results in the exclusion of 1,296,481 students, which constitute 48.9% of the data. Therefore, an analysis based on a dataset without ethnically homogeneous schools is unlikely to be a representation of differences between schools. However, the magnitude of the results presented here may be skewed towards the scores obtained by students in these schools.

The results presented here are based on data for all the students, schools and LAs described in chapter 4. Appendix A.4.2 shows that the overall results presented here also hold for the sample restricted to schools with all ethnic groups and for the schools that are observed during all years in the data, with some differences in the estimated magnitudes.

This section is divided into three parts. First, section 6.5.1 presents the trends in the overall gap, derived from the total-effect models; second, section 6.5.2 presents the components of the gap, derived from the outcome and mediation models; finally section 6.5.3 combines these findings to decompose the ethnic achievement gap into its within-and between-school and LA components.

6.5.1 The Overall Ethnic Achievement Gap

The data analysis chapter (section 4.3) showed that there are differences in the distribution of maths test scores among ethnic groups. Figure 6.5 focuses on the differences in the mean test scores between White and minority students, i.e. the ethnic achievement gap for each minority group. The pooled comparison of White versus all minority groups ('All minorities' in the figures) serves as a reference to contrast the results with Sánchez-Jabba (2011)'s work, that does not differentiate among minority groups.

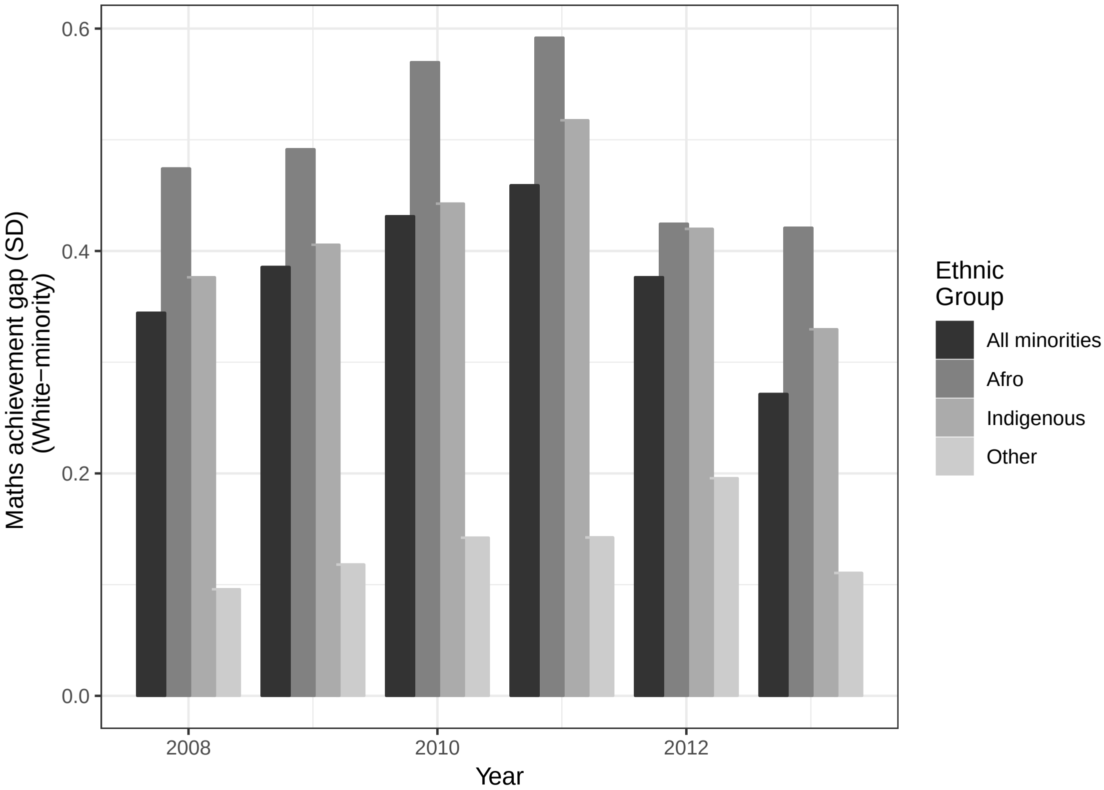

Figure 6.5: Ethnic achievement gap in maths test scores for each ethnic minority and for all minorities combined

Figure 6.5 shows that the difference in mean test scores between White and minority students increases from 0.34 SD in 2008 to 0.46 SD in 2011 to then decrease to 0.27 SD in 2013. The ethnic achievement gaps for Afrocolombian and Indigenous students follow a similar trend, as Afrocolombian students, the group with the consistently widest achievement gap, score 0.47 SD, 0.59 SD and 0.42 SD lower than White students, and Indigenous students score 0.38 SD, 0.52 SD and 0.33 SD lower than White students during 2008, 2011 and 2013, respectively. The trend is slightly different for the group of other minority students as the achievement gap for this group ranges between 0.1 SD in 2008 and 0.2 SD in 2012, which is consistently and considerably lower than the gap for Afrocolombian and Indigenous students. These results highlight the differences among ethnic minority groups, which means it is worth analysing them separately instead of pooling them together as a single group, as in Sánchez-Jabba (2011).

6.5.2 Components of the Gap

Sections 6.3 and 6.4 showed that the overall ethnic achievement gap is the result of combining the within-school gap β^{W}, the school and LA contextual effects of ethnicity β^{CS} and β^{CL} (which together create between-school and LA ethnic differences in academic achievement β^{BS} and β^{BL}) and a series of segregation indices ω. This section studies these components before section 6.5.3 combines them to decompose the ethnic achievement gap into its within- and between-school/within LA and between-LA components.

6.5.2.1 Within-School Gap and School and Local Authority Contextual Effects

As explained in section 6.4, the within-school ethnic gap and the school and LA contextual effects of ethnicity can be estimated using contextual effects models, which are the outcome model in the mediation analysis framework used to decompose the gap. Figure 6.6 shows the estimation results for two different outcome models. The left-hand-side column presents the estimated parameters for model (6.12), which excludes the LA proportion of minority students, and thus ignores this contextual effect. The results on the right-hand side column are for model (8.1), which includes this term.

As expected, both models provide the same estimate of the within-school ethnic gap, which is relatively stable over time, unlike the overall gap. When comparing White and all minorities grouped together, White students score on average between 0.05 SD (in 2009 and 2010) and 0.07 SD (in 2011 and 2013) higher than their minority peers attending the same school. However, examining the within-school gap for each minority group reveals further differences between ethnic groups. Afrocolombian students score around 0.08 SD lower than their White peers attending the same school, while Indigenous and other minority students score around 0.06 SD and 0.03 SD lower than their White peers. These results show how the pooled gap (comparing White and all ethnic minorities as a single group) hides further information about each minority group and the importance

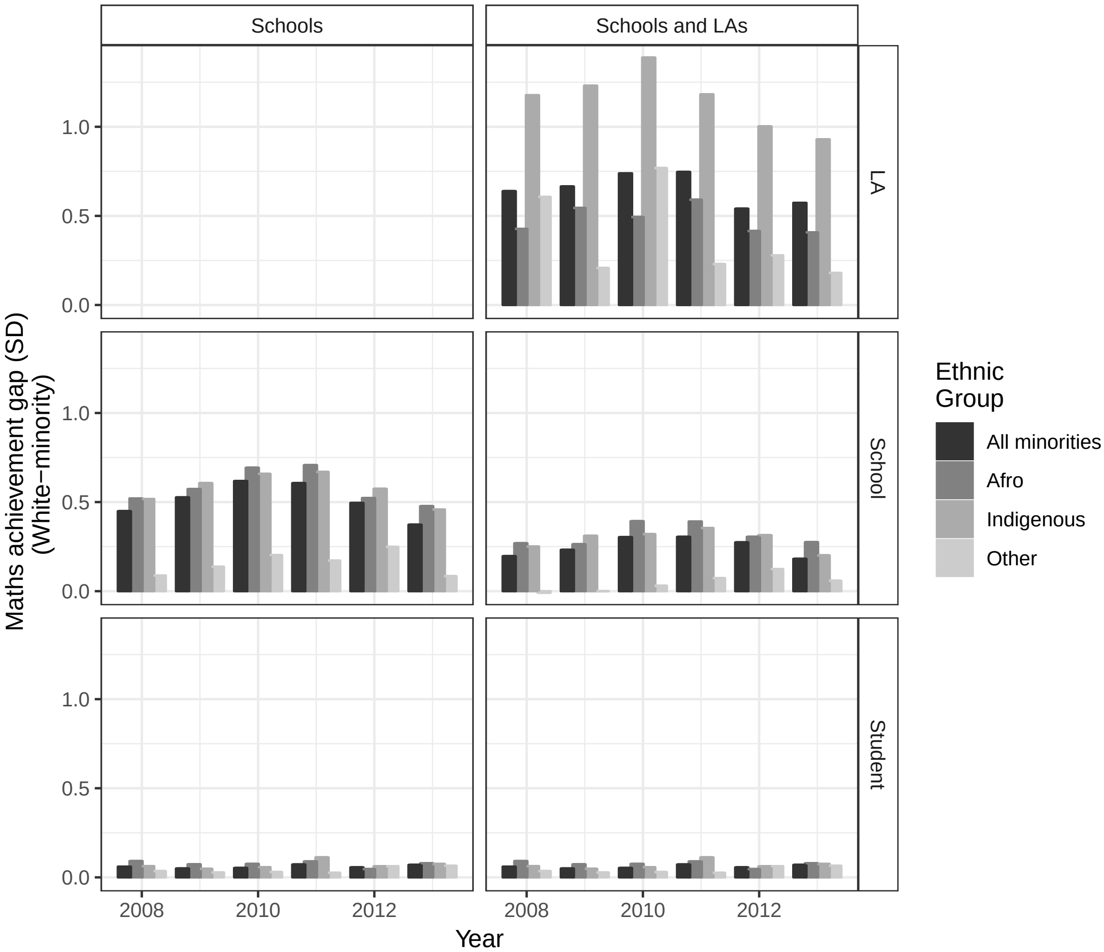

Figure 6.6: Within-school gaps and school and LA contextual effect of each minority group for models including and excluding the LA contextual effect of ethnicity

of considering them separately for a more relevant policy formulation.

Similarly, when analysing the school contextual effect of ethnicity, focusing on the White-minority comparison masks the particularities of each ethnic group, as shown in Figure 6.6. In particular, the pooled White-minority comparison for the school contextual effect of ethnicity shows an increase from 0.45 SD in 2008 to 0.62 SD in 2010 and then a drop to 0.37 SD in 2013, as shown in the left-hand-side column of Figure 6.6. The school contextual effects of Afrocolombian and Indigenous students follow a similar trend to the overall ethnic achievement gap, increasing from 0.52 SD for both ethnic groups in 2008 to 0.71 SD and 0.67 SD for Afrocolombian and Indigenous students, respectively, in 2011 and then decreasing to 0.48 SD and 0.46 SD for these two ethnic groups. In turn, the school contextual effect of the group of other minority students does not follow a clear pattern, ranging between 0.08 SD in 2013 and 0.25 SD in 2012. This implies that students in schools with higher proportions of minority students tend to obtain lower scores, independently of their own ethnicity. Nonetheless, these results may also reflect effects at higher levels, such as the LA-level.

The right-hand-side column of Figure 6.6 explores this possibility, which shows that the LAs' ethnic composition explains more than half of these reported school contextual effects. Namely, schools with a larger proportion of minority students tend to be located in LAs with a higher proportion of minority students. Therefore, considering the role of LAs allows separating these two effects and thus improves the pertinence of evidence for policymaking. As a result, it is shown that an additional ten percentage points in the school proportion of Afrocolombian or Indigenous students is linked to a ($\frac{\beta^{CSg}}{10}$) decrease in the student's test scores between 0.05 SD (in 2008 and 2013) and 0.07 SD (in 2010 and 2011), independently of their ethnicity (β^{SC}). In turn, a ten percentage points increase in the school proportion of other minority students is associated with a decline of between 0.01 SD (in 2008, 2009 and 2013) and 0.02 SD (in 2010, 2011 and 2012) in the mean students' test scores of students of all ethnicities.

Figure 6.6 also shows that the LA contextual effect of ethnicity is also different for each ethnic group. For Afrocolombian and the group of other minority students, this effect is similar to the school contextual effect, varying between 0.41 SD (in 2013) and 0.59 SD (in 2011) for Afrocolombian students and between 0.18 SD (in 2013) and 0.77 SD (in 2010) for the group of other minority students. In turn, the LA contextual effect of Indigenous students is larger than the school contextual effect of this ethnic group, varying between 0.93 SD (in 2013) and 1.39 SD (in 2010). This means that an increase of ten percentage points in the LA proportion of Indigenous students is associated with an average decrease of 0.12 SD in the student's maths test scores, independently of their ethnicity and the ethnic composition of their schools.

These results show that the within-school gaps are small compared to the contextual effects and the overall gaps, which implies that it can be expected that differences between schools and LAs play a more critical role in explaining the overall achievement gap. However, the contribution of each of these differences to the overall ethnic achievement gaps for each ethnic group depends on the segregation indices ω, which are further examined in the next section.

6.5.2.2 Segregation Indices

As section 6.3.1 explained, the overall ethnic achievement gap is a weighted sum of the within-school gap and the school and LA contextual effects of ethnicity, examined in the previous section. The weights of this sum are the segregation indicators ω_Y^X, which are the difference in the school (or LA) average proportion of students of the ethnic group X, for students of group Y and White students. Therefore, when $\omega_Y^X = 0$, it means that the school (LA) proportion of students of the group X for students of the group Y is the same as the school (LA) proportion of students of the group X for White students. These indicators can be estimated using the mediator models in sections 6.3 and 6.4. Figure 6.7 depicts these indicators, representing the group X by the columns and the group Y by the shade of the bars.

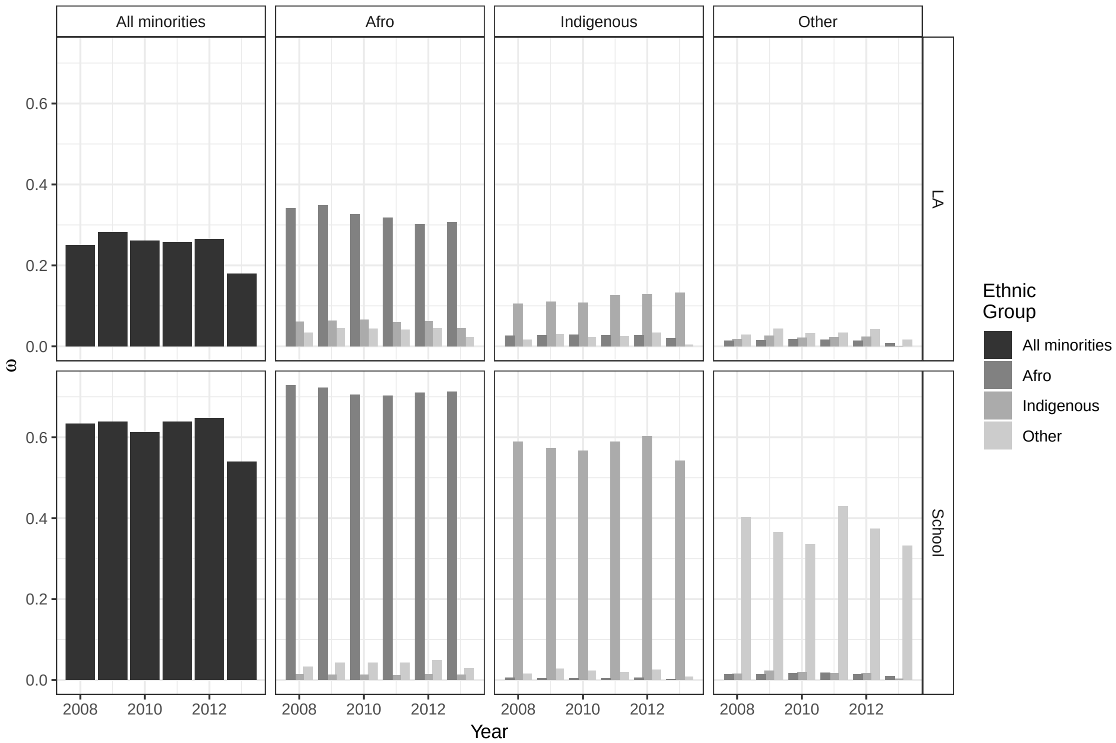

Figure 6.7: Segregation indices ω_Y^X in the achievement gap decomposition

The first column of Figure 6.7 shows the pooled comparison of White and all minority students. It shows that consistently minority students live in LAs with around 25.9 percentage points more minority students, on average, in comparison to the LAs in which White students live; and attend schools with around 63.6 percentage points more minority students than the schools White students attend, on average.

The next columns of Figure 6.7 show that all minority groups live in LAs with a higher average proportion of students of their own ethnic group in comparison to White students. Students also attend schools in which a higher proportion of students belong to their same ethnic group. In particular, Afrocolombian students live in LAs with around 32.3 percentage points more Afrocolombian students than the LAs in which White students live, on average. Besides, Afrocolombian students attend schools with around 71.1 percentage points more Afrocolombian students than the schools White students attend, on average[10].

In contrast, Afrocolombian and Indigenous students tend to attend schools in which the proportion of Indigenous and Afrocolombian students, respectively, is not much higher than that for the schools White students attend. That is, Afrocolombian students attend schools with 0.4 percentage points more Indigenous students, on average, in comparison to White students. In turn, Indigenous students attend schools with 1.3 percentage points more Afrocolombian students than the schools White students attend[11].

Given the strong contextual effects of ethnicity shown in section 6.5.2.1 and these high levels of segregation, it is expected that differences between schools and LAs mostly explain the overall ethnic achievement gap. This is further examined in the next section.

6.5.3 Ethnic Achievement Gap Decomposition in Practice

This section presents the results for the ethnic achievement gap decomposition into its within- and between-school and LA parts. As argued in section 6.3.2, in its simplest within- and between-school form, such decomposition is given by

$$\beta = \left(1 - \omega^S\right) \beta^W + \omega^S \left(\beta^W + \beta^{CS}\right)$$

where $\left(1 - \omega^S\right) \beta^W$ is the within-school component of the gap, which is linked to the within-school gap, and $\omega^S \left(\beta^W + \beta^{CS}\right)$ is the between-school component of the gap, which is linked to the between-school gap. Section 6.4 extended this decomposition to

[10] These numbers decrease to 31.8 and 65.3, respectively, when the sample is restricted to those students attending mixed schools.

[11] When focusing on ethnically mixed schools, Afrocolombian students attend schools with 0.1 percentage points more Indigenous students than the schools White students attend. Besides, the difference between the proportion of Afrocolombian students in schools for Indigenous and White students reduces to 0.7 percentage points.

also consider differences at the LA level and multiple minority groups. Therefore, this section is further divided into two sections. Section 6.5.3.1 presents the results for the White-minority comparison, while section 6.5.3.2 presents the results of the decomposition for multiple minority groups.

6.5.3.1 White-Minority Gap Decomposition

Figure 6.8 shows the proportion of the achievement gap between White and minority students that can be attributed to the within-school and the between-school gap (left-hand side panel), and within-school, between-school/within-LA and between-LA gap (right-hand-side panel). As expected, both decompositions result in the same within-school component of the gap, which varies between 4.7% (0.02 SD) of the overall gap in 2009 and 11.87% (0.03 SD) of the overall gap in 2013. These changes in the percentual contribution of within-school differences is mainly driven by changes in the overall gap, since, expressed as standard deviations, the within-school component of the gap only varies between 0.02 SD and 0.03 SD, reflecting the relatively stable nature of the within-school gap and the segregation indicator.

Considering additional levels is equivalent to further decomposing the between-school component of the gap. When the LA-level is not considered, differences between-schools account for between 88.1% (0.24 SD in 2013) and 95.3% (0.37 SD in 2009) of the White-minority achievement gap. Using this result to recommend focusing on school-level policies or holding schools accountable for ethnic inequality could be misleading in this context, as considering differences between LAs reveals that between-school differences account for only between 26% (0.1 SD in 2009) and 33.5% (0.13 SD in 2012) of this gap. In turn, differences between LAs account for between 54.8% (0.15 SD in 2013) and 69.4% (0.27 SD in 2009) of the overall White-minority achievement gap.

Given these results, extending the decomposition methodology to consider additional levels of the school system reveals that a more important focus of policy intervention is the differences between LAs. The policy relevance of this evidence can also be improved by additionally considering differences between minority groups, as each of them may require a different kind of intervention if the relative importance of within- and between-school and LA differences varies among groups. This is further examined in the next section.

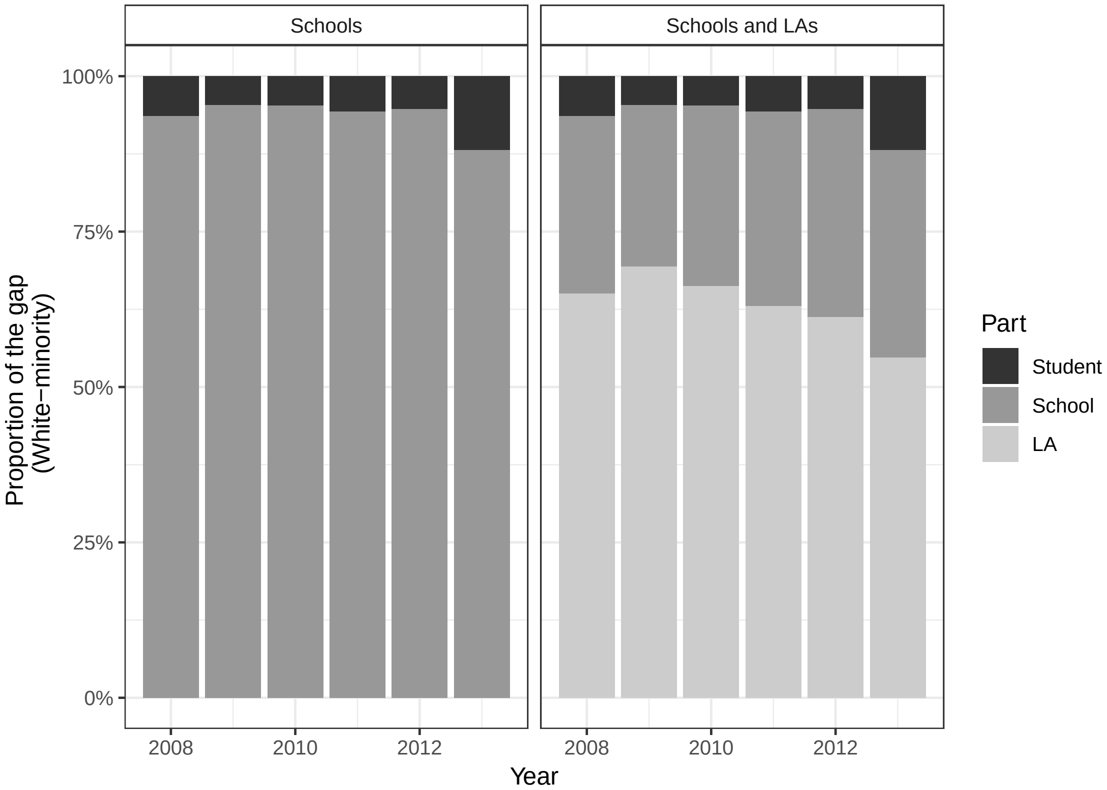

Figure 6.8: Decomposition of the overall achievement gap between White and minority students

6.5.3.2 Considering Differences Among Minority Groups

As explained in section 6.4.2, decomposing the ethnic achievement gaps for each ethnic group requires considering the role of other ethnic groups, which adds a level of complexity to the analysis. This section starts by analysing the simpler within- and between-school decomposition and then incorporates the LA-level into the analysis.

Figure 6.9 shows the results of decomposing the overall ethnic achievement gaps for Afrocolombian, Indigenous and other minority students into their within- and between-school components. The within-school component of the White-Afrocolombian achievement gap accounts for between 3.2% (0.01 SD in 2012) and 5.4% (0.02 SD in 2013) of the overall gap. Similarly, it accounts for between 5% (0.02 SD in 2009) and 10.4% (0.03 SD in 2013) of the White-Indigenous overall achievement gap. In contrast, the within-school component accounts for between 10.3% (0.01 SD in 2011) and 39.6% (0.04 SD in 2013) of the achievement gap for other minority students.

These differences between ethnic groups also translate into the between-school com-

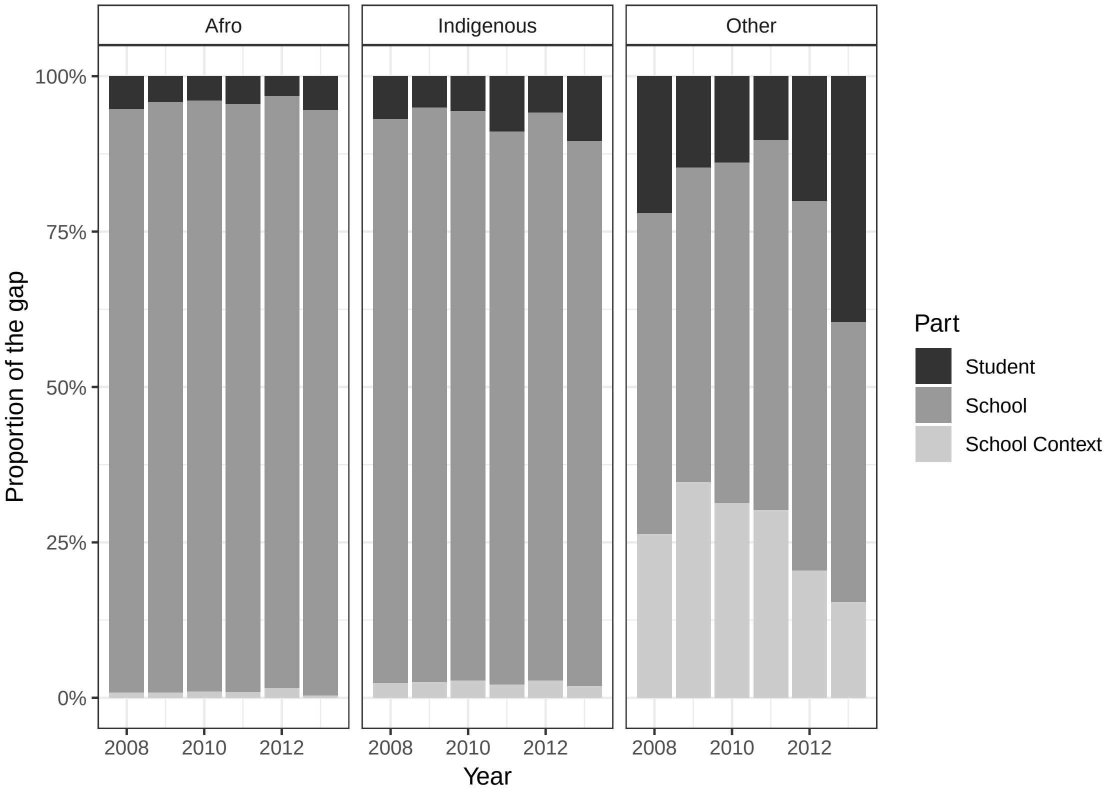

Figure 6.9: Decomposition of the overall achievement gap between White and Afrocolombian, Indigenous and other minority students into its within- and between-school components

ponent of the gaps, which is associated with the between-school ethnic achievement gap (school), and a component linked to the school contextual effect of other ethnic groups (school context). As shown in Figure 6.9, the between-school component is much larger than the school-context component, despite the school contextual effects of all ethnic groups being large (as discussed in section 6.5.2.1). For example, between 93.9% (0.45 SD in 2008) and 95.3% (0.4 SD in 2012) of the achievement gap for Afrocolombian students can be attributed to the between-school gap for Afrocolombian students, while only between 0.4% (0.002 SD in 2013) and 1.6% (0.007 SD in 2012) of this gap can be attributed to the school contextual effect of other ethnic groups.

This result is explained because Afrocolombian (and, similarly, Indigenous) students are less disproportionately exposed to students of other ethnicities as they are to students of their same ethnic group; as shown in section 6.5.2.2, the segregation indices are much larger with respect to their own ethnic group than with respect to other groups.

In turn, the group of other minority students are more disproportionately exposed to Afrocolombian and Indigenous students, and therefore the proportion of the gap for the group of other minority students that is explained by the school context is between 15.4% (0.02 SD in 2013) and 34.7% (0.04 SD in 2009) of the overall White-other gap. The between-school component of the gap for the group of other minority students ranges between 45% (0.05 SD in 2013) and 59.5% (0.12 SD in 2012) of the overall gap.

As with the White-minority gap decomposition in section 6.5.3.1, ignoring the role of LAs could lead to less effective policy recommendations. Therefore, Figure 6.10 shows the achievement gap decomposition into its within-, between-school/within-LA and between-LA components for each minority group. This figure reveals that differences between LAs are generally much more important than differences between schools, especially for Afrocolombian and Indigenous students. For Afrocolombian students, for example, the between-school component of the gap accounts for between 25.6% (0.13 SD in 2009) and 34.2% (0.14 SD in 2013) of the overall White-Afrocolombian achievement gap, whereas the between-LA gap component ranges between 54.7% (0.23 SD in 2012) and 62.5% (0.31 SD in 2009) of this gap. This component of the gap is more variable for the group of other minority students, ranging between 4.7% (0.01 SD in 2013) and 19.1% (0.03 SD in 2010) of the overall White-other gap.

Figure 6.10 also shows that the context component is more important at the LA level than at the school level for all minorities, but it is more important for the group of other minority students, given the large LA contextual effects of ethnicity and that there is more exposure to other minority groups at this level, as shown in section 6.5.2. For instance while the LA contextual effect (of Afrocolombian and other minority students) explains between 5.6% (0.02 SD in 2013) and 11.2% (0.05 SD in 2010) of the White-Indigenous overall gap, the LA contextual effect of Afrocolombian and Indigenous students explains between 12.1% (0.01 SD in 2013) and 52.4% (0.06 SD in 2009) of the White-other gap. In turn, the school-context component constitutes between 0.8% (0.003 SD in 2009) and 1.6% (0.007 SD in 2012) of the gap for Indigenous students, but between 8.5% (0.009 SD in 2013) and 16.9% (0.024 SD in 2010) of the White-other gap.

These results reflect the finding in section 6.5.2.2 that students in the 'other' minority group are much more disproportionately exposed to Afrocolombian and Indigenous students than these later groups to any other ethnic minority. However, there is no clear pattern in the components of the gap to explain the increase between 2008 and 2011 and

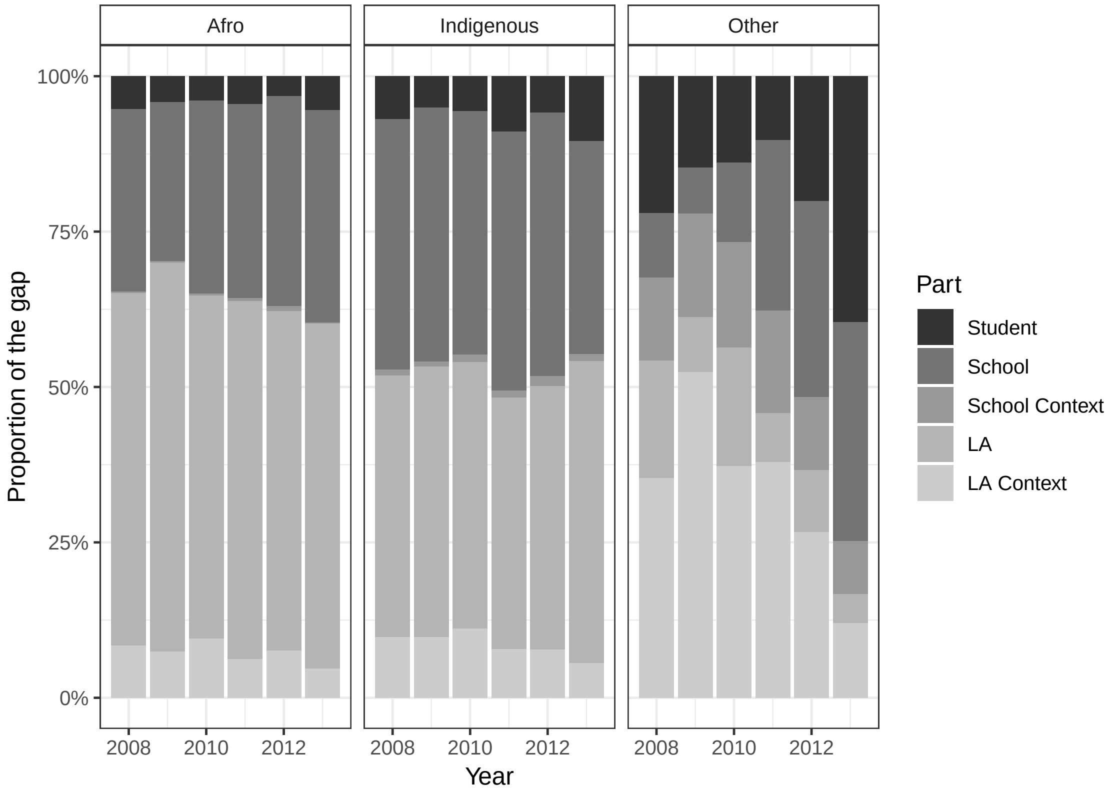

Figure 6.10: Decomposition of the overall achievement gap between White and Afro-colombian, Indigenous and other minority students into its within-, between-school and between-LA components

later drop in the overall ethnic achievement gap, which may be explained by a more complex combination of changes in the components of the gap or by changes in the way the test favours specific ethnic groups over time. These results are further discussed in the next section.

6.6 Discussion

6.6.1 Summary of Findings

6.6.1.1 Methodological Findings

Section 6.3 showed that the ethnic achievement gap had been decomposed in three different ways. Cook and Evans (2000) and Fryer and Levitt (2004, 2006) decomposed the Black-White achievement gap into the within-school gap (which they interpreted as the contribution of within-school differences to the overall gap) and the effect of segregation (which they interpreted as the contribution of between-school differences). Hanushek and Rivkin (2006) decomposed the gap into a part that is attributable to the within-school gap and a part that is attributed to the between-school gap, as it has been shown outside the ethnic-achievement-gap literature (e.g. Causa & Chapuis, 2011; O. D. Duncan et al., 1961; Willms, 2010). Finally, Reardon (2008) and Page et al. (2008) argued that there is a third component of the gap that cannot be unambiguously attributed to either within- or between-school differences. This chapter showed that this component is attributed to differences in the student intake and appears after separating the contribution of the between-school gap into a part that arises from the differences within-schools and a second part that results from the school contextual effect of ethnicity, which Reardon (2008) and Page et al. (2008) claimed was unambiguously attributed to between-school differences.

Each of these decompositions is useful to inform different research questions and policy decisions. For example, if the debate is about school segregation, Cook and Evans (2000)'s and Fryer and Levitt (2004, 2006)'s decomposition provides a more direct way to analyse its potential effects on the achievement gap. If, in turn, the interest is in what the mechanisms behind the between-school component of the gap are, Reardon (2008)'s and Page et al. (2008)'s decomposition can be used. If, as in this chapter, the focus is on the within-school and between school components of the gap, Hanushek and Rivkin (2006)'s decomposition is appropriate. However, this debate was limited in that it did not appropriately consider the role of multiple ethnic groups (beyond the binary White-minority comparison) or additional levels of the school system (such as LAs), which restricts the kind of policy recommendations that follow from the application of these methods.

Hence, this chapter proposed using a mediation model in which the school's ethnic composition mediates the ethnic achievement gap as a tool to decompose the overall ethnic achievement gap into the contributions of within- and between-school differences. Translating the gap decomposition debate into the mediation framework allows extending the current decomposition methodology to analyse more complex problems and overcoming the limitations mentioned above. These extensions incorporate multiple mediators and categorical variables to consider the role of LAs and multiple ethnic groups. It is not argued that mediation analysis is the only method that can be used to extend this methodology. Directly finding the mathematical formulation is another option, for example. Nonetheless, translating the achievement gap decomposition into a mediation analysis frameworks unlocks many additional potential extensions, besides the ones considered in this chapter, as further discussed in section 6.6.2.

The role of between-LA differences (or the equivalent in the US, school districts) has only be considered using Cook and Evans (2000)'s method of including school- (in this case, LA) fixed effects into a Oaxaca (1973) decomposition. Arteaga and Glewwe (2019) proposed using this method to examine the extent to which the gap between Indigenous and non-Indigenous students could be attributed to community effects in Peru. The use of the Oaxaca (1973) decomposition implies a different kind of decomposition than the one explored in this chapter. Essentially, Arteaga and Glewwe (2019) decomposed the gap in four components: the role of observable student characteristics ('endowment effect'), differences in the way these characteristics affect Indigenous and non-Indigenous students ('relative impact effect'), differences in where Indigenous and non-Indigenous students live ('community sorting effect') and differences in the way in which communities affect Indigenous and non-Indigenous students ('heterogeneity in community impact effect'). Similarly, Dustmann et al. (2010) considered different ethnic groups by using Cook and Evans (2000)'s decomposition method, which, as discussed above, does not decompose the gap into its within- and between-school components, but into the within-school gap and the effect of segregation. In turn, Quinn (2015b) used Reardon (2008)'s approach to decompose the Black-White gap in the US and included dummy variables for other ethnic groups to ensure they included all possible observations while comparing Black and White students (instead of Black and non-Black students). Nonetheless, this approach did not examine the role that other minority students play in explaining the Black-White achievement gap, which is also a component of the Black-White gap, as explained in section 6.4.2.

The application of these extensions to the Colombian context presents new evidence of the ethnic achievement gaps and allows for a richer debate about them, as discussed in the next section.

6.6.1.2 Substantive Findings

The overall ethnic achievement gaps (the average differences in test scores between White and minority students) in Colombia vary between 0.42 SD and 0.59 SD for Afrocolombian, 0.33 SD and 0.52 SD for Indigenous students in 2013 and 2011, respectively, and 0.1 SD and 0.2 SD in 2008 and 2012 for other minority students, which shows the importance of studying each ethnic group in separate instead of using Sánchez-Jabba (2011)'s approach of pooling together all ethnic minorities. These gaps are in general not as wide as those reported in the US, which reached 1.2 SD for 8th grade (age 13) Black students in 1971 or the more recent 0.89 SD gap in 2011. Nonetheless, the achievement gap for Afrocolombian students in 2011 reached similar levels to the 0.58 SD reported for 8th grade Black students in 1988 in the US (Page et al., 2008). The achievement gap for Indigenous students in Colombia is also similar to the 0.39 SD gap for 6th grade (age 11/12) Indigenous students in Guatemala in 2001 (McEwan & Trowbridge, 2007) and the 0.4 SD gap for 8th grade (age 12/13) students in Chile in 1997, but higher than the 0.34 SD gap for 6th grade (age 11) Indigenous students in Bolivia in 1997 (McEwan, 2004). In England, the ethnic achievement gaps vary between ethnic groups much more than in Colombia. In the former, Chinese students score on average 0.88 SD higher than White British students, while Black Caribbean and Roma students score on average 0.25 SD and 1.48 SD, respectively, lower than White British in 2016 (Leckie & Goldstein, 2019). Since the ethnic achievement gaps in England narrow as students progress through schooling, the gaps for 11th grade (age 16/17) students in Colombia are more similar to those between White British Year 6 (age 11) and Black Caribbean (0.45 SD) or Black other (0.37 SD) students in 2008 (Strand, 2014b). Each of the school systems listed here have distinctive features, such as the extent of academic tracking or how market-oriented the organisation of schools is. Identifying how or why these features result in wider or narrower ethnic achievement gaps is out of the scope of this thesis, but pointing out these differences informs future research.

There is also a pattern on the overall gaps for Afrocolombian and Indigenous students increasing between 2008 and 2011 and then dropping in 2012 and 2013, which is robust to different sample selection criteria, as shown in appendix A.4. This pattern may arise from four different sources[12]. First, changes in the components of the gap (within-school gaps, contextual effects and segregation) although, as further discussed below, these components do not follow the same pattern. Second, changes in other disparities among ethnic groups, such as an initial worsening and later improvement in socioeco-

[12] As explained in section 3.3.2.1, the resource-allocation rules changed from 2011 to not only consider the geographical dispersion of the population but also consider how socioeconomically disadvantaged LAs are. Nonetheless, the effects of such a change in policy are expected to be detected in the long term and not immediately.

nomic inequalities, although the data chapter mentioned that the distribution of these characteristics is stable over time. Third, as mentioned in the context chapter (section 3.3.3), from 2007 and until 2011 LAs with a high proportion of minority students in the south-west of the country transitioned from a type-B (September to June) to a type-A (February to November) school year. The transition process might have had detrimental effects on the students' performance, as it was carried out by gradually shortening the school year of state schools in these LAs, but there is no evidence of the impact this transition had on the students' achievement. Finally, these trends may reflect differences in the SABER 11 exam, that initially may have disproportionately favour White students in comparison to ethnic minorities and then adjusted these differences after 2011. A study of the 2007 version of the exam concluded that the test does not favour any particular ethnic group (Cuevas Mendoza, 2013), but this has not been replicated for other versions of the exam that, as mentioned in section 4.2.1, has changed since 2007. The available data do not allow distinguishing between these explanations.

As discussed before, the overall gap is the result of combining the within-school gap with the school and LA contextual effect of ethnicity, according to the level of segregation. Authors do generally not report these parameters for a model without controls. This implies that, for example, the reported estimates of the within-school gap result of comparing White and minority students with the same set of observable characteristics within the same school, which are explored in chapter 8, instead of the coarser estimates in this chapter that compare White and minority students regardless of their characteristics. The discussion that follows focuses on the few studies that report the unconditional ethnic within-school gap and contextual effects (or between-school gaps), while chapter 8 contrasts the findings of the conditional within-school gaps and contextual effects in Colombia with those for other school systems around the world.

The within-school gaps of around 0.08 SD, 0.06 SD and 0.03 SD for Afrocolombian, Indigenous and other minority students, respectively, are smaller than the 0.45 SD Black-White within-school gap for 5th grade (age 10/11) students in the US in 2004 or the 0.75 SD Black-White gap for 8th grade (age 13/14) students in 2011 (Bohrnstedt et al., 2015; Hanushek & Rivkin, 2006; Reardon, 2008), but it is similar to the 0.07 SD Black-Caribbean-White within-school gap for Year 11 (age 16) students in England in 2007 (Dustmann et al., 2010). The narrower gaps in this context may imply that schools provide a more uniform experience for all students in Colombia, in comparison to these other countries, which may be aid by the lack of tiering or ability grouping. Nonetheless, this may also imply that students within schools are also similar in terms of other characteristics (such as their SES) independently of their ethnicity, and therefore a similar average achievement could be expected.

The school contextual effect for Afrocolombian (around 0.55 SD) and Indigenous students (around 0.59 SD) without considering the role of LAs (as the papers in the literature do), is much larger than the 0.13 SD and 0.33 SD school contextual effect of fifth (age 10, in 2004) and eighth grade (age 13/14, in 2011) Black students in the US, respectively (Bohrnstedt et al., 2015; Hanushek & Rivkin, 2006)[13]. These figures are, however, more similar to the more variable school contextual effect of the group of other minority students of between 0.08 SD and 0.25 SD. As discussed in the literature review chapter (section 2.5.3) this may be the result of pure peer effects, teachers adjusting their expectations and teaching methods according to the ethnic composition of their schools, as found in Colombia (Valoyes Chavez, 2015), or differences in school quality.

This chapter showed that considering the role of LAs allows one to estimate the school contextual effect of ethnicity within LAs, which results in a school contextual effect of around half these magnitudes. Since none of the studies discussed in this chapter estimated the unconditional LA (or school district) contextual effect of ethnicity, further comparisons with the international context are not possible.

Finally, segregation is comparatively high in Colombia. As shown in section 6.5.2.2, Afrocolombian students attend schools with an average of 71.1 percentage points more Afrocolombian students than the schools White students attend. Similarly, Indigenous students attend schools with an average of 58.1 percentage points more Indigenous students than the schools White students attend. Such levels of school segregation exceed those for 8th grade (age 13/14) Black students in the US in 2011 (39 percentage points) and even in 1971 (52 percentage points) (Bohrnstedt et al., 2015; Page et al., 2008), and the segregation levels for Year 11 (age 16) minority students in England in 2007 (at most 54.3 percentage points) (Dustmann et al., 2010). These high levels of segregation may be the result of school and neighbourhood selection processes, which may be based on ethnicity, but also on other characteristics, such as SES (Barbary & Olivier, 2003; Villegas

[13]Notice that since other authors in the debate and applications included school-fixed effects in the model, they did not estimate the school contextual effect of ethnicity (e.g. Cook & Evans, 2000; Dustmann et al., 2010; McEwan & Trowbridge, 2007).

García & López Quiroz, 2011) and the operation of market-like mechanisms of school competition in a school system that favours private over state schools (Murillo, 2016). Segregation between LAs is likely to obey historical processes such as the colonial heritage that left Afrocolombian groups living near the ports that were used for slavery and in places where slaves escaped to, while Indigenous groups mainly live in their ancestral territories (Manrique et al., 2003; Ministry of Culture, 2010).

The combination of high segregation with small within-school gaps and large school and contextual effects results in a small within-school component of the gap, which accounts for up to 5.4%, 10.4% and 39.6% of the overall ethnic achievement gap for Afrocolombian, Indigenous and other minority students, respectively. The contribution of within-school differences to the overall gap are larger for the group of other minority students because they are less segregated across schools and LAs than students of other ethnic minorities. However, all these contributions of within-school differences are much smaller than the 13.4% and 19.9% contribution of the within-school gaps to the Black-White achievement gap for 5th grade (age 10) students in 2004 in the US estimated by Hanushek and Rivkin (2006) and Reardon (2008), respectively[14], and than the 51.6% contribution of within-school differences for the Black-White gap for 8th grade (age 13/14) students in 2011 estimated by Bohrnstedt et al. (2015)[15]. Hanushek and Rivkin (2006) attributed the remaining 86.6% to differences between-schools, while Reardon (2008) and Bohrnstedt et al. (2015) further decomposed this component into a 39.5% and 16.1% attributed to differences between schools and 40.6% and 32.3% to the 'ambiguous' component of the gap, respectively. In Colombia, given the small size of the within-school gap in comparison to the contextual effect of ethnicity, this 'ambiguous' component is much smaller than in the US. In contrast, further dividing the between-school component to consider the role of LAs reveals a large between-LA component of the gap, which accounts for at least 54.7%, 40.5% and 4.7% of the overall gap for Afrocolombian, Indigenous and other minority students respectively, which illustrates the importance of taking this higher level into account given its policy implications,

[14]Although Hanushek and Rivkin (2006) and Reardon (2008) used the same database (ECLS-K), their sample sizes differed, which may explain the difference in these results.

[15]Please notice that other authors replicated Cook and Evans (2000)'s methodology and therefore argued that the contribution of within-school differences was even higher, as they interpreted the part of the between-school component that is due to differences in student intake between-schools as part of the within-school gap contribution.

discussed in section 6.6.2.

6.6.2 Implications for Policy

The ethnic achievement gap decomposition in this chapter allows a further understanding of the nature of the ethnic achievement gap and how policies can affect it. As shown in section 6.3, the ethnic achievement gap depends on three basic components: the within-school gap, the (school and LA) contextual effect of ethnicity and the level of segregation, measured as the differential exposure to other ethnic groups. Modifying each of these components represent different paths for policy.

Reducing the within-school achievement gap, i.e. the systematic advantage of White students over their minority peers in the same school, would lead to a reduction of differences both within- and between-schools with a different ethnic composition. If, for example, minority students tend to be more socioeconomically disadvantaged than their peers attending the same schools (as is the case for Indigenous students), such policies would ensure disadvantaged students enjoy the same experiences that advantaged students do.

Policies that aim to reduce the contextual effects of ethnicity would lead to a reduction of disparities in the mean test scores between schools and LAs with a different ethnic composition. As with the within-school gap, if schools and LAs with a higher proportion of minority students are also socioeconomically disadvantaged, an example of a policy to eliminate the contextual effects of ethnicity would be to ensure that these schools and LAs have access to the same resources as socioeconomically advantaged schools and LAs. However, as pointed out before, if the within-school (student-level) differences persist, differences between schools and LAs would remain.

Finally, (de)segregation policies change the importance of within- and between-school and LA gaps in determining the overall gap. Unless the within-school gap and the contextual effects have different signs[16], the within-school gap is smaller than the between-school and between-LA gap. In that case, reducing segregation narrows the overall gap because it increases the importance of the (smaller) within-school compo-

[16]For example, if within-schools minority students have higher average scores than their White peers, but students in schools and LAs with larger proportions of minority students achieve lower scores on average. Chapter 9 shows that minority students score higher than their White peers within some schools, but the magnitude of these differences was not big enough to counteract the negative school and LA contextual effects.

nent of the gap. However, this narrowing of the gap does not necessarily imply an improvement of differences within or between schools and LAs with a different ethnic composition, unless (de)segregation changes the within-school gap or the school and LAs contextual effects of ethnicity. This is potentially the case, as Card and Rothstein (2007) found that more segregated cities in the US had a wider within-city Black-White achievement gap, which implies that desegregation would also be a mechanism to reduce between-school/within-city differences.

In Colombia, the within-school gap is small, and there are large negative effects of studying in schools and LAs with higher proportions of minority students. Therefore, reducing the high levels of segregation would narrow the overall ethnic achievement gaps because the smaller within-school differences would become more important for the overall gap, assuming that the components of the gap are unaffected by such a change. If, as in the US, desegregating schools narrows the within- and between-school (and LA) gap, there would be an additional incentive for a desegregation policy. Nonetheless, the feasibility of such a policy requires further consideration of the ways to preserve the right of ethnic minorities to remain and rule in their territories. In any case, reducing within- and between-school disparities should be a priority of public policy, given their large magnitudes. Such a policy also requires a different approach for each ethnic group, as this chapter shows that the importance of each of the components of the gap varies by ethnic group. Chapters 8 and 9 discuss in more detail what are the considerations that possible interventions may take into account, as these may require changes in some school and LA characteristics, but also on their composition, depending on the ethnic groups and subgroups of students, schools and LAs.

These policy recommendations assume that the gaps are reduced by improving the academic achievement of ethnic minorities, which would raise overall achievement. Policies that aim to equalise students regardless of the level of academic achievement may risk harming overall achievement. Nonetheless, policies that only aim to improve academic achievement without considering how these improvements are distributed among ethnic groups risk increasing ethnic inequality.

6.6.3 Limitations

The limitations of this chapter arise from its methodological and substantive parts. The methodological discussion about the ethnic gap decomposition is tailored to its substantive application to the Colombian context, which implies that it only considers three levels that are observed on the data: students, schools and LAs and three ethnic minority groups. This restricts the detailed discussion of additional levels (such as cohorts and classrooms) and ethnic groups (which in England, for example, are usually reported using many more categories). Nonetheless, the decomposition method can be generalised to these additional levels and groups, as explained in appendix A.4.1. The use of the mediation analysis framework should also facilitate the extension of the decomposition to consider more complex nested structures, such as cross-classified or multiple membership structures that are not considered in this chapter. If these structures provide a better representation of the ethnic achievement gaps, by recognising the role of neighbourhoods or multiple schools that students have attended, for example, the decomposition in this chapter may overstate the contribution of schools and LAs to the overall ethnic achievement gap (Browne, Goldstein, & Rasbash, 2001; Goldstein, 1994; Leckie, 2013a, 2013b). Nonetheless, the data restrictions do not allow evaluating the impact of these alternatives.

Another limitation of the method for decomposing the ethnic achievement gap, as presented here, is that it does not allow for any interactions. Therefore, the models assume that the contextual effect each minority group is the same for all ethnic groups or, equivalently, that the within-school gaps are the same regardless of the proportion of minority students in the schools and LAs. In the context of the Oaxaca (1973) decomposition, which Hou (2014) showed also fits under the mediation analysis framework, this is resolved by estimating separate equations for each group, which is an approach that can be further explored to expand the decomposition in this chapter. In the context of this thesis, however, chapter 9 shows that there are no interactions between the within-school gaps and the school or LA ethnic composition, which implies that this extension is not crucial for decomposing the ethnic achievement gaps in Colombia. Similarly, the school and LA contextual effects of ethnicity are assumed to be linear[17], which is a standard assumption in this literature (e.g. Bohrnstedt et al., 2015; Quinn, 2015b; Reardon, 2008), but it could be evaluated in future research.

This chapter did not study the uncertainty of the components of the gap. For example, it did not examine whether the 3.2% of the overall gap for Afrocolombian students

[17]This is, that the change in maths test scores is the same for an increase from 10% to 20% in the proportion of minority students than for an increase from 80% to 90% of this value.

that is attributed to within-school differences is the result of random chance (whether it is statistically different from zero). This type of analysis requires deriving the formulation for the standard errors of the components of the gap or using non-parametric techniques such as bootstrapping. This is out the scope of this thesis, but in this chapter, repeating the analysis for different cohorts provides evidence that the findings are a feature of the ethnic achievement gaps and not the results of random shocks, even though their uncertainty is not measured.

The substantive application in this chapter shares the limitations derived from the conceptual framework of this thesis, which focuses on ethnic achievement gaps and no other aspect of schooling or academic achievement, as discussed in section 1.4. It also shares the limitations derived from the data availability, as discussed in section 4.6, which prevents any longitudinal analysis or the analysis of different school levels, regarding the substantive application.

Additionally, the estimated within-school gaps in this chapter result from comparing White and (Afrocolombian, Indigenous, other or pooled) minority students attending the same schools, which ensures that the differences in test scores are net of the effect of school- and LA-level variables. Nonetheless, these differences do not separate socioeconomically advantaged and disadvantaged students or students with other characteristics that favour maths achievement. Since White and minority students attending the same schools also have different characteristics, such as SES, as shown in the data chapter, the within-school gap is not the pure effect of ethnicity. Similarly, the school and LA contextual effects of ethnicity do not only reflect the effect of the school ethnic composition, if there is any, but also the influence of other school and LA characteristics, including their composition. Chapters 7 and 8 explore the role of these characteristics.

The focus in this chapter (and in chapters 7 and 8) was on the ethnic achievement gap, in isolation from its relationship with academic achievement. The study of the ethnic achievement gap cannot wholly ignore possible consequences for overall achievement, as both increasing achievement and reducing inequality are desirable social outcomes. The links between these two are explored in chapter 9.

Finally, this chapter assumed that the ethnic achievement gaps and their components are the same for all type of students, schools and LAs. For example, implying that gaps are the same in LAs with good and poor resource management. Chapter 9 explores if this is the case.

6.6.4 Implications for Future Research

As discussed in the previous section, there are several limitations of this chapter. Some of them are addressed in the following research chapters, but others need to be addressed in future research as they are out of the scope of this thesis. In particular, this chapter presents the first study to propose using the mediation analysis framework as a tool to decompose the ethnic achievement gap into its within-and between-school components and illustrates two examples for which using this framework allows extending the current decomposition method. However, there are many other potential extensions, including the consideration of more complex nesting structures than the hierarchical one analysed in this thesis. Another area for methodological research is to develop tools for statistical inference for the components of the gap. Similarly, these extensions can be applied as an exploratory step of the analysis of ethnic achievement gaps that allows gaining more insight into the type of processes that could explain this phenomenon, but also in applications of a variety of disciplines that analyse nesting structures. Examples include the gender pay gap, ethnic inequality in health outcomes or differences in political orientation.

Substantively, further research is required to support more relevant policies that tackle the ethnic achievement gap. Such research should consider differences among minority groups and prioritise understanding the role that LAs and schools play in ethnic inequality. An important aspect is understanding whether and how segregation affects the within-school gaps and the school and LA contextual effects of ethnicity, as well as other potential outcomes such as the well-being of students and inter-group behaviour (McKeown, Cairns, Stringer, & Rae, 2012). This knowledge is essential to understand if reducing segregation is a useful tool for reducing the ethnic inequality or if the ethnic achievement gap only would change because the within- and between-school achievement gaps receive a different weight.

6.7 Summary

Following its aims, this chapter:

1. Used mediation analysis for the first time as a tool to explain and extend the ethnic-achievement-gap-decomposition debate to consider differences at three (or more) levels of the education system and for multiple ethnic minority groups. This

translation allowed arguing that:

(a) The ethnic achievement gap is the result of combining the within-school gap with the school and LA contextual effect of ethnicity according to the segregation levels (measured as exposure).

(b) Each of the methodological approaches proposed in the achievement gap decomposition debate sheds different lights about the ethnic achievement gap and can be useful when answering different questions. Nonetheless, a within-school, between-school interpretation of the components of the gap is in line with Hanushek and Rivkin (2006).

(c) Translating the debate into a mediation analysis framework unlocks the potential for further extensions that are already available within this framework. This not only includes adding ethnic groups and levels to the decomposition, as shown in this chapter, but also extensions such as considering latent variables and measurement error.

2. Showed that the between-school gaps and the level of segregation (measured as exposure) in Colombia are high in comparison to other countries, while the within-school and overall gaps are relatively small.

3. Illustrated the importance of extending this decomposition methodology by showing that the most important component of the ethnic achievement gaps for 11th grade (age 16/17) students in Colombia are the between-LA components. Additionally, there are differences among ethnic groups, with differences between LAs being more important for the gap between White and Afrocolombian students than between White and Indigenous students.

These findings provide a roadmap for the remainder of this thesis, showing the importance of studying differences at various levels of the school system (students, schools and LAs) for each ethnic minority group. The next chapter explores one of the most common explanations for ethnic achievement gaps, SES, and how its operationalisation may affect the findings.

7 How do the Different Index Construction Techniques for Operationalising Socio-Economic Status used in the Literature Affect the Estimates of the Conditional Ethnic Achievement Gaps and their Components at each Level of the Education System?

7.1 Introduction

Chapter 6 focused on the unconditional achievement gap; the test score differences between White and minority students that are observed across the school system. As discussed in the theoretical framework in section 2.7, these differences reflect how White and minority students are different in many aspects. One of these aspects is socio-economic status (SES), which has the particularity of being operationalised in many ways in the literature. The simplest approach would be to add SES to the models explored in chapter 6. An immediate difficulty is that as most of the SES-related variables in the SABER 11 dataset are recorded as categorical, there is a large number of parameters to be estimated and the variables tend to be highly correlated, which may lead to multicollinearity problems, including inflated standard errors, unstable estimated parameters and problematic interpretations of the estimation results (Johnston et al., 2018). One alternative to tackle this challenge is to create a composite indicator to measure SES, which may also be easier to communicate to a wider audience. Nonetheless, this may be at the cost of losing valuable information or incorporating distortions into the analysis when summarising the raw indicators.

In the social-science literature, SES is conceptualised as the position of individuals within society, according to their ability to access material and non-material resources, with education, occupation and income (known as 'the big three') as its main elements

(Bornstein et al., 2003; Buchmann, 2002; Finch & Hoehn, 1951; Jeynes, 2002; Sirin, 2005; White, 1982). Nevertheless, one aspect of the operationalisation of SES without consensus yet is the use of composite indicators. There are two main gaps in the advice provided to researchers in this regard. The first one is whether it is preferable to use a set of variables or a composite indicator to control for SES. The second one is how different index-construction methods compare to each other. Besides, it is unclear what are the index-creation methods that are typically used to measure SES in research about achievement gaps.

The operationalisation of SES can potentially affect the estimated conditional ethnic achievement gap, depending on the relationships between the SES indicator and achievement and SES and ethnicity, as discussed in the methodology chapter (section 5.3). Besides, depending on how the SES indicator clusters within schools, it may affect how the conditional achievement gap is divided into its within-school gap and school and local authority (LA) contextual effects components, explored in chapter 8. Furthermore, the clustering patterns of the SES indicator may affect the proportion of explained between-school variance in maths achievement, which also has consequences on the estimation of random-slope models estimated in chapter 9, which allow for different within-school achievement gaps for each school.

Using the inclusion of 'the big three' SES variables as a reference (education, occupation and income), this chapter examines how the conclusions about the conditional ethnic achievement gaps change when controlling for SES by using different composite indicators. These indices differ in the method that is used to create them, which, based on the literature review presented in section 7.2.2, include factor analysis (FA), principal component analysis (PCA), the sum and average of standardised variables, the sum of unstandardised variables, the Hollingshead (2011) four-factor index and prediction from linear regression.

As further discussed in section 7.2.1, most of the studies that have previously analysed the operationalisation of SES have focused on either theoretical reasoning, the size of the effect of SES on achievement or the predictive power of the SES variables. However, these are not the only factors that may affect the conclusions of the analysis of the ethnic achievement gap (or other gaps in general). This chapter advances the discussion by considering the elements that affect the estimation of conditional ethnic achievement gaps and by introducing a multilevel perspective into the analysis. While there is a substantive interest in the differences in SES among ethnic groups and their effect on the achievement gap, this chapter focuses the discussion on the measurement of SES. Chapter 8 presents the substantive analysis of how SES affects the ethnic achievement gaps.

This chapter is organised as follows:

1. Section 7.2 reviews the relevant literature, identifying the index-construction techniques that are used in the study of achievement gaps in different countries.

2. Section 7.3 briefly presents the methods and data used to create and analyse the SES composite indicators, explaining how their inclusion affects the estimation of the ethnic achievement gap.

3. Section 7.4.1 creates the SES composite indicators using different index-construction techniques. At this stage, the only concern is understanding how the indices differ in the way they represent SES, how they correlate with observed variables and other indicators, and how they are distributed among ethnic groups.

4. Section 7.4.2 examines if the indices vary in their power to predict academic achievement, measured by maths test scores, and whether it changes by ethnicity.

5. Section 7.4.3 evaluates how the indices cluster within schools to understand if the index-creation process generates any distortion in the original clustering pattern.

6. Section 7.4.4 inspects how the ethnic achievement gaps conditional on SES change according to the SES indicator, exploring how the models estimated in chapter 6 change when adding SES as a control variable.

7. Section 7.5 discusses these results.

8. Section 7.6 summarises the contributions of the chapter.

7.2 Measuring SES when Studying Achievement Gaps: Review and Current Practice

Many researchers have studied the role that SES plays in academic achievement and other social and health outcomes. For this reason, several reviews have been conducted to understand the importance of SES and, to a lesser extent, how practitioners operationalise this variable. Nonetheless, these reviews do not usually discuss the methods used to create composite indicators.

This literature review consists of two parts. Section 7.2.1 presents a review of reviews about the effect of SES on academic achievement and other outputs, aiming to understand how researchers operationalise SES broadly. Section 7.2.2 presents a systematic literature review that provides more insights into the use of composite indicators for measuring SES by focusing on the use of composite indicators in studies about gaps in academic achievement.

7.2.1 Review of Reviews

White (1982) conducted a meta-analytic review of 101 documents, mainly from the US and Canada, that explored the relationship between SES and academic achievement. He found a low (0.22) median correlation between achievement and SES, while reporting that correlations varied between 0.1 and 0.8. White (1982) also showed that one reason for studies reporting such dissimilar values for this correlation was the variety of variables used to represent SES, as the correlation between achievement and home resources was 0.57, while the correlation between achievement and education was 0.18. Besides, White (1982) found that education, occupation and income were traditionally used. Nonetheless, White (1982) did not further discuss the details of the operationalisation of these variables, beyond the use of one or more variables simultaneously.

Sirin (2005) replicated White (1982)'s meta-analysis for the period between 1990 and 2000, with very similar results. He also found that reported correlations between academic achievement and SES varied widely (between 0.005 and 0.77) with a low (0.24) median correlation. Also similarly, Sirin (2005) found that even though education, occupation and income were commonly used to represent SES, the correlation estimates varied according to the selected variable, with measures of home resources producing a mean correlation of 0.51, while parental occupation produced a mean correlation of 0.28. However, the discussion was again limited to what variables were used and not their operationalisation.

In contrast with these two studies, which focused on the size of the correlation between SES and academic achievement, other studies have explored how SES is conceptualised and its broad implications for data collection and analysis. Buchmann (2002) reviewed the way family background, including SES, was measured in comparative educational research, focusing on the Programme for International Student Assessment (PISA) and the International Association for the Evaluation of Educational Achievement (IEA) studies -like the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS)-. Her review showed a wide diversity in the ways SES is operationalised around the world when studying the relationship between SES and academic achievement, with some studies using indices and others using a different set of variables (generally including a measure of education, some proxy for wealth and with some exceptions, occupation). Although extensive, this review did not provide information about the methods that were used to compute the indices in the reviewed studies. Besides, since Buchmann (2002) focused on international educational research, her recommendations focused on how to collect comparable information in future international surveys. The document did not provide recommendations for researchers using existing databases for secondary analysis.

Similarly, Yang has extensively studied the measurement of SES in comparative research. Her work has mainly tried to understand what aspects of SES affect academic achievement and the comparability of SES indicators. Yang (2003) and Yang and Gustafsson (2004) investigated how to use IEA's studies' home possessions data to measure SES at the student and school levels, whether the possessions data could be represented by the same number of dimensions in different countries, and whether the relationship between SES and maths and reading achievement was the same around the world. These studies used multilevel structural equation modelling to conclude that in most countries, SES could be represented using two dimensions (cultural and material resources) at the student-level and one dimension at the school-level, although there was some variation between countries. In the countries for which only one dimension was enough to represent SES, there was a positive relationship between that dimension and reading achievement. For those countries with a cultural and material dimension of SES, the cultural dimension was related to reading achievement, while the material dimension of SES was not. In another study, Yang Hansen and Munck (2012) argued that it was necessary to study the psychometric properties of SES since it was more common to use more than one variable to measure this construct. In this occasion, SES was operationalised through multilevel latent class analysis, which allowed the authors to identify three SES classes at the student level using the Swedish PIRLS data. Yang Hansen and

Munck (2012) also found a positive relationship between SES and reading achievement. However, Yang has not compared the methods in these papers (factor analysis measurement models in Yang (2003) and Yang and Gustafsson (2004) and latent class analysis in Yang Hansen and Munck (2012)). That is, the question of whether there is an empirical reason to use a particular method to measure SES as a continuous factor or as a small number of homogeneous classes is still unanswered.

Ensminger and Fothergill (2003) reviewed 471 papers on child and adolescent development published between 1991 and 2000 in the USA, with a closer look at the operationalisation of SES. They found that mother's education was the most frequently used SES indicator and that the use of composite indices was uncommon. They reported that, of the few articles using composite indicators, the majority used the Hollingshead (2011) four-factor index of social status. Ensminger and Fothergill (2003) recognised that further research is required to understand how composite measures are used in this kind of studies and how their inclusion impacts on the conclusions of the studies. This request has not been answered in following specialised reports.

For example, the APA Task Force on Socioeconomic Status (2007) reviewed the conceptual discussion and importance of SES for psychological research. They pointed out that most research used education, occupation and income to measure access to resources. They argued that each of these domains had a different causal pathway, and thus, the authors recommended measuring SES as a separate set of variables instead of using composite indicators. Therefore, the APA Task Force on Socioeconomic Status (2007) did not provide any recommendations regarding the creation of composite indicators. This position is understandable when the piece of research aims to identify the causal mechanisms behind SES. However, studies that consider SES may have a different focus, as is the case in this thesis, which focuses on ethnicity.

Cowan et al. (2012) also reviewed the definition of SES to provide recommendations about its measurement for the National Assessment of Educational Progress (NAEP). They conceptualised SES as "one's access to financial, social, and human capital resources" (Cowan et al., 2012, p.14). They made a clear distinction between core and expanded measures of SES, as well as contextual/explanatory variables of achievement. The core measures are educational attainment, occupational status and family income and other measures of resources. According to them, these are essential to any measure of SES. The expanded measures of SES could complement the core and include indicators of the neighbourhood and school SES. These are different from contextual measures such as perceptions and psychological processes, which do not constitute SES but extra predictors of test scores.

In contrast to the APA Task Force on Socioeconomic Status (2007), Cowan et al. (2012) recommended using composite indicators for measuring SES instead of using (a set of) single variables, arguing that the former are easier to interpret and report, less prone to reporting error and capture information on multiple observed variables, in comparison with the later. They underscored the distinction between formative (the indicators cause SES) and reflective (SES causes the indicators) measurement models and argued that formative models are more in tune with the definition of SES. However, they recognised that deciding how to weight the indicators requires further research, considering the literature and the quality of the data.

7.2.1.1 Focus on the Specific Research Contexts

A smaller group of studies have attempted a comparison between ways of operationalising SES in specific contexts, most of them are concerned about the comparability of SES across countries. Most of these studies explored the consequences of changing the variable definitions and using composite indicators instead of sets of variables but do not explore different methods for creating them.

In the UK, Willms (1992, ch. 5) studied the best predictors of school performance in the 7th year of primary and 4th year of secondary school in Scotland. The predictors he tested included family background (parental occupation and education, number of parents and number of siblings), sex and prior achievement. After examining the R^2 of models including and excluding family background and prior achievement, Willms (1992) concluded that prior achievement was a more important predictor of achievement than family background. However, using prior achievement is not an option in this thesis, as discussed in the data chapter (section 4.6).

More recently, Ilie, Sutherland, and Vignoles (2017) questioned whether Free School Meal (FSM) status[1] accurately identified socio-economically disadvantaged children and whether other variables (parental occupation, parental attainment, house ownership,

[1]Free School Meal (FSM) status is commonly used in educational research in the UK as a measure of SES; see for example Borooah and Knox (2015); Burgess and Briggs (2010); Leckie and Goldstein (2019); Rasbash et al. (2010); Strand (1999, 2016).

family structure, income and neighbourhood child deprivation) were better predictors of academic achievement. They concluded that, even though FSM status did not accurately identify all children in deprived households, the predictive power gains did not offset the costs of gathering information on other variables. Ilie et al. (2017) also found that area-level predictors are poor predictors of academic achievement. The study, however, only tested the power of individual variables and all variables included at the same time, without attempting any combination of them as a compound index.

Also in the UK context, Lenkeit, Caro, and Strand (2015) tackled a different problem by questioning if SES and economic, social and cultural capital had a shared meaning for families of different ethnic background. They found that all these constructs have a different meaning for students of different ethnic background, arguing that the educational, occupational and cultural status in the country of origin did not translate well into the English context. They also found that the association between SES and language achievement was very similar between ethnic groups, but this did not hold for the measures of economic, cultural and social capital. As the focus of Lenkeit et al. (2015) was on measurement invariance and on the difference between SES and economic, cultural and social capital, they operationalised all these measures using exploratory structural equation modelling and different approaches to operationalising these variables were not discussed, beyond mentioning a better fit of this method in comparison to using FA.

Kahl and Davis (1955) compared FA and cluster analysis for combining 19 different measures of education, occupation, income and social status perception. They concluded that the measures were highly correlated and that these two methods were "complementary ways of reducing the [inter-correlations] matrix to simpler and more understandable form" (Kahl & Davis, 1955, p. 325). However, they compared the methods only in terms of the information they could provide about SES, but not of how they influenced the analysis results when applied to a specific research problem.

In the context of comparative research, Rutkowski and Rutkowski (2013) examined the reliability and cultural comparability of the PISA home possessions index and recommended researchers to compute their own SES indicators, arguing that the reliability of the index was not acceptable for some countries. However, they did not offer a reason to prefer using composite indicators instead of operationalising SES as a set of variables. In turn, G. N. Marks (2011) contrasted these two approaches to investigate how conclusions regarding socioeconomic inequality across PISA 2000 participants would change.

He reached similar (although not equal) conclusions using both approaches. Additionally, he experimented with different variable definitions and their inclusion in the creation of a composite indicator using sheaf coefficients[2]. G. N. Marks (2011) recommended using composite indicators but excluding measures of material and cultural resources, arguing that it is not possible to know if they are a cause or a consequence of SES. Nonetheless, G. N. Marks (2011) did not evaluate if his conclusions changed when using different methods to create the composite indicators.

Caro and Cortés (2012) referred to studies of the IEA when proposing a procedure to calculate and validate an SES indicator using the PIRLS 2006 data. This procedure mimicked the creation of the PISA ESCS Index, using PCA, and complemented it by proposing to evaluate the validity and reliability of the index. To achieve this, they proposed studying the degree of correlation between the input variables, checking that the weights of the indicators are the same for all countries and that the indices are positively related to achievement, and calculating three different measures of reliability (the Cronbach's alpha, the beta coefficient and the greater lower bound coefficient). However, Caro and Cortés (2012) did not consider deriving the SES index using other procedures beyond PCA. They also focused on illustrating how to create the index, but they did not compare the approach of using an index versus including the set of variables in the model.

In the ethnic achievement gaps literature, the effect of different operationalisations of SES on the conclusions about the influence of SES on ethnic inequality has not been studied. In health research, Braveman et al. (2001) examined whether using education and income and their level of measurement (categorical or continuous) made a difference in studies about ethnic disparities in maternal and infant health. They concluded that the operationalisation of SES made a difference in the results of the role of ethnicity in these health indicators and that this difference varied by ethnicity, variable of interest and SES control. For example, the conditional (on income) odds of delivering a low birth-weight baby for a Latina were 67% higher than the odds for a European/Middle Eastern woman if income was measured as a categorical variable, while the same figure increased to 107% if income was measured as a continuous variable. The authors

[2]Sheaf coefficients are a method for measuring latent variables that assigns weights to the observed variables (e.g. education, occupation and income) in such a way that the latent variable is the best possible predictor of the outcome variable (e.g. maths test scores) (Whitt, 1986).

recommended researchers to be transparent about the effect that choosing different SES indicators may have on their results and to ground their choices on theoretical considerations. Braveman et al. (2001) compared different operationalisations of income and education, but they did not evaluate the possibility of combining these variables using composite indicators. Besides, all the health outcomes they examined were measured as binary variables, and therefore, their analysis was limited to binary logistic models.

For studies using the Colombian National Assessment Institute (Icfes) data as well as other administrative datasets in Colombia and Latin America, there are two tendencies: including a set of variables (e.g., education, occupation, income, ownership of durable goods) into their regression models to operationalise SES (Noe et al., 2005; Sánchez-Jabba, 2011; Viáfara López & Serna Alvarado, 2015), or processing the large number of available variables using PCA or FA to construct a SES indicator (Orjuela, 2013; Piñeros Jiménez & Rodríguez, 1998; C. Rangel & Lleras, 2010; Stein et al., 2005; Vinha et al., 2016). What papers in these two tendencies, and in general papers around the world have in common, is that they are not transparent about the effects that selecting a different way to operationalise SES may have on their results. That is, papers that use PCA or FA do not discuss how their results would change had they operationalised SES as a set of variables, and papers that use a set of variables to represent SES do not interrogate the possibility of using a composite indicator or a different set of variables.

Even though SES and its measurement have been broadly discussed across disciplines, over time and around the world, the literature is still far away from a consensus in terms of how to operationalise SES. There is a gap in the literature about the use of composite indicators, as it is still unclear how they are used and whether there is an empirical reason to prefer a particular method for creating such indicators in studies about achievement gaps. The first step, presented in the next section, is understanding how researchers operationalise SES in studies about academic achievement gaps.

7.2.2 Systematic Literature Review

This section presents the results of a systematic literature review that focuses on the operationalisation of SES in studies about gaps in academic achievement, the focus of this thesis. As such, it is much more specific than the review of reviews presented in section 7.2.1. The tool to conduct this review was the Scopus database, which compiles around seventy million peer-reviewed abstracts and citations of over twenty thousand serial titles, from over five thousand publishers in many research fields, including social science, and around the world. The serial titles in Scopus have passed the review of an independent advisory board, which decides whether or not to include a title according to its policy, content, standing, regularity and online availability (Elsevier B.V., 2017). The above implies that Scopus does not include records on all possible research publications, but it does provide a comprehensive view of quality-assessed sources around the world. Even though it probably disproportionally excludes local literature (written in local languages without abstracts in the English language), Scopus includes documents from a broader variety of regions than alternative databases (Elsevier B.V., 2017).

The following sections describe the findings of that search, focusing on the use of composite indicators and the methods that are adopted for creating such indices. These findings will guide the development of this research chapter, which interrogates the reason for selecting one way of operationalising SES over the others in the context of the study of ethnic achievement gaps in Colombia.

7.2.2.1 The Search Strategy

The Scopus search, carried out on 29 January 2018, included the following fields:

Abstract and key-words “socioeconomic status” & achievement or test scores & gap

Source type Journals

Document type Articles

Languages English, Spanish and Portuguese

These fields restrict the search to articles in academic journals that mention socioeconomic status in the context of (any) gaps and achievement or test scores. The search also filters articles written in languages other than English, Spanish or Portuguese in an attempt to include papers from as many countries that could be a relevant comparison for Colombia, including Brazil. Then, this review of research practice relies on the assumption that the methods used by the documents found through this search are representative of the methods that researchers around the world currently use. Based on papers outside this particular search, such as those reviewed in chapter 2, this seems to be a reasonable assumption.

The search returned 202 documents, which included 39 documents that were not empirical studies of a gap in academic achievement considering socioeconomic status. These documents included, for example, nine articles about health, nine theoretical papers and five documents based on qualitative research. Another ten papers conducted their analysis at the school, local authority or higher level of analysis, and not at the student-level.

The remaining 153 documents were published between 1984 and 2018 in 109 different journals. These used data from international studies and 31 different individual countries using 60 different databases. Eighty seven of the papers focused on the USA and fifteen analysed more than one country simultaneously. Altogether, the articles presented research from industrialised countries, such as the UK, Germany and Japan, and developing countries as diverse as Israel, Guatemala and Cambodia.

Most (81) papers measured SES because this was the main variable of interest, while 56 documents used SES as a control variable in regression analysis and 14 articles measured SES to describe their sample. One document used SES to split its sample.

7.2.2.2 Composite Indicators versus Individual Variables

The number of papers that used composite indices and individual variables to measure SES was almost even. Seventy two of the reviewed papers used some sort of composite indicator to measure SES, while 76 used between 1 and 7 individual variables to represent SES. It was unclear how the remaining five papers measured this variable.

Figure 7.1 shows the number of articles that were found by their year of publication and whether or not they use a composite indicator to measure SES. The figure shows that the number of published papers about gaps in academic achievement using SES has increased over time. Nonetheless, there is no obvious pattern in the use of composite indicators to measure SES over time. In 2012, 60% of the articles used composite indicators, but only 20% of the papers published in 2010 did so.

Both, papers with SES as the main variable and articles using SES as a control variable used composite indicators to measure SES. Forty four of the 81 documents focusing on SES used composite indicators, and 28 of the 56 papers using SES as a control variable measured this variable using a composite index.

The use of composite indicators seems to be linked to its availability in the database used in the study. For example, only six of the 40 articles that collected their own data

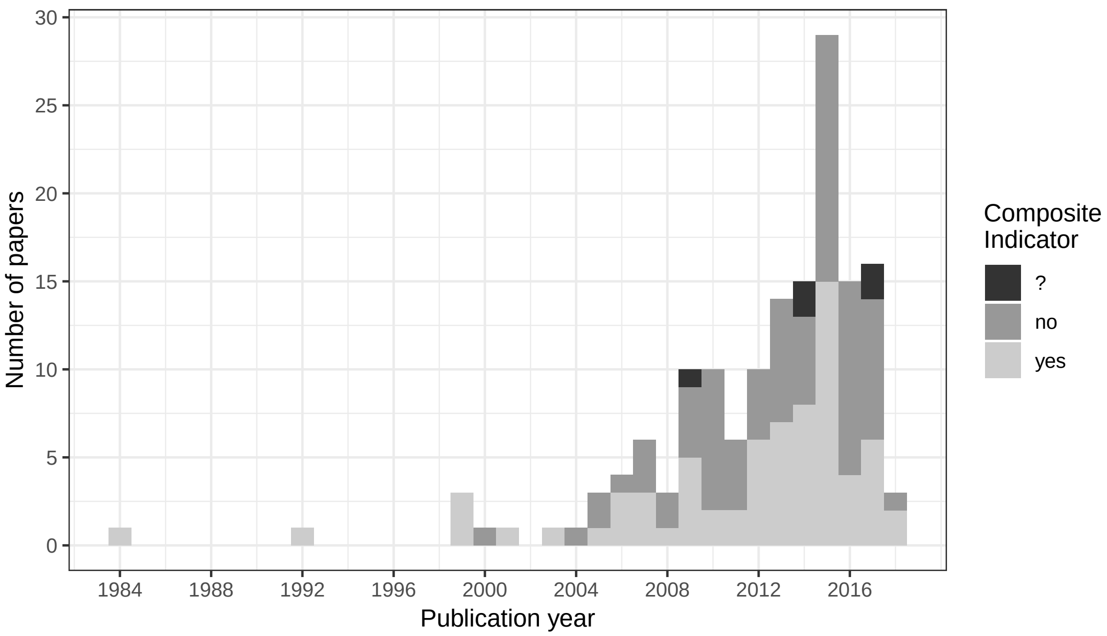

Figure 7.1: Number of papers by year of publication and use of composite indicators to measure SES

used a composite indicator to measure SES, while 66 of the 113 papers using secondary data did so. In particular, 33 of the 41 papers using databases from the National Centre for Education Statistics in the USA (NCES)[3] used a composite indicator and only one claimed to have created their own. Similarly, nine of the twelve papers based on the PISA databases used composite indices of SES and only four articles created their own.

7.2.2.3 Commonly Used Variables

Consistently with what has been previously pointed out in reviews about the measurement of SES in studies of academic achievement (Buchmann, 2002; Sirin, 2005; White, 1982), documents prevalently used education, occupation and income to measure SES, although free-school meal eligibility was often used as a sole indicator of SES, particularly in the UK and the USA.

Most (39) of the 76 studies that did not use composite indicators measured SES using only one variable. Twelve of them used FSM eligibility status (or its equivalent; reduced lunch eligibility in the USA, for example). Ten of these papers opted for education and

[3]The papers in this review used the National Education Longitudinal Study, NELS 1988; the Early Childhood Longitudinal Study, ECLS, the Education Longitudinal Study of 2002; ELS:2002 and the High School & Beyond, HS&B 1980.

eight for income as a measure of SES. Only six documents used occupation as their only measure of SES. Education and income (each of them used in 28 articles) were most often used when papers included more than one measure of SES without combining them into a composite indicator. Though parental education is often measured as the highest parents' educational attainment.

In the 21 cases in which the articles created their own SES composite indices, authors used between 1 and 12 variables to create such indicators, with most (17) of them using less than five variables. Fifteen papers used education, 12 used occupation and 11 used income to create their indicators. Most papers with five variables used both parent's education and occupation and family/household income, while those with three variables usually employed mother's education, father's occupation and family income. Articles with larger numbers of variables often collected information on assets in the household.

7.2.2.4 Commonly Used Composite Indicators

The sections above have shown that a substantial proportion of papers used a SES composite indicator, created either by the authors (most likely) using data collected by themselves or by the administrators of the secondary data that authors analysed. Consequently, the discussion in this section focuses on the methods that have been used in the reviewed papers for creating composite indicators, and on the indicators provided in different datasets.

The 21 papers that created their own SES indicators used factor analysis (FA) (four documents), principal component analysis (PCA) (five articles), the average of standardised variables (three papers), the sum of standardised variables (two papers), the sum of raw variables (two papers), the Hollingshead (2011) four factor index of social status (one paper) and prediction from linear regression (one paper). It is unclear how two of the documents combined the variables to create their composite indices.

As shown in section 7.2.2.2, 31 documents used a SES composite index that someone else had already created using the secondary database they analysed. Most of these employed the NCES' longitudinal studies (20 articles) and the PISA's databases (seven papers). The remaining articles employed the PIRLS, the Growth and Efficiency Measures of Schools database (in Hebrew, *Meitzav*) 2008 and 2009 for Hebrew-language schools in Israel, the Institute of Nutrition of Central America and Panama (INCAP) Human Capital Study 2002-2004 study in Guatemala, and the SiBO longitudinal study

Table 7.1: Methods for creating SES indices used by papers in the review

Method	Own Index	Others' Index
Factor analysis (FA)	4	
Principal component analysis (PCA)	5	7 - PISA, Caro and Cortés (2012), INCAP
Average of standardised variables	3	21 - NCES, SiBO
Sum of standardised variables	2	
Sum of unstandardised variables	2	
Hollingshead (2011)	1	
Prediction from linear regression	1	

in Belgium.

The paper using the PIRLS's database employed the index computed by Caro and Cortés (2012), which is based on PCA. It was not possible to find the method that the Israeli Ministry of Education used for creating the social deprivation index employed in the article using the *Meitzav* database. The methods to create the composite indicators in the reviewed papers are summarised in Table 7.1. The second column of this table shows the number of papers that created their own index, using each method in the first column of the table. The third column shows the database(s) (or paper) that employed each method.

7.2.2.5 Summary

The previously described search in Scopus, although not exhaustive, provided an overview of how researchers operationalise SES in peer-reviewed papers about academic achievement gaps around the world.

One striking finding is that it is not always clear how researchers measured SES in their study. Those who clearly stated how they operationalised SES could be divided between those who used a set of separated variables (or only one variable) and those who used composite indicators. The use of composite indices seems to be linked to their availability in the dataset under study, which is the case of some large-scale assessments. In contrast, documents that also involved data collection exercises did not commonly use composite indicators. In any case, consistent with what seems to be a consensus in the literature, most studies understood SES as a combination of education, occupation and income/wealth (Buchmann, 2002; Ensminger & Fothergill, 2003; Sirin, 2005; White, 1982), although FSM eligibility was also used.

Regarding the methods for creating composite indicators, there seems to be much more variety in indices created by the authors than in indices created by database administrators, as shown in Table 7.1. Importantly, although almost invariably education and occupation were collected as categorical variables, none of the methods in Table 7.1 are especially suitable for the analysis of this kind of data. In most cases, the main variables undergo some transformation before being combined using the methods in Table 7.1. For example, before using PCA to create the PISA index of economic, social and cultural status, occupational categories are transformed into occupational prestige scales using Ganzeboom, De Graaf, and Treiman (1992)'s socio-economic index of occupational status, and educational attainment is transformed into years of education. Whether these transformations affect the results is further explored in this chapter.

These studies reflect the consensus about the variables that constitute SES. At the same time, they reflect the lack of consensus about the best way to operationalise these variables, especially when it comes to combining them as an index.

7.3 Methods and Data

As followed from the literature review in section 7.2.2, the authors who measure SES using composite indicators, typically use FA, PCA, the sum or average standardised or raw variables, linear prediction or the Hollingshead (2011) four-factor index to measure SES. Therefore, this chapter compares these methods against each other and against raw variables in terms of: the assumptions they make about the way in which SES and the observable variables interact, how they vary in their capacity to predict maths achievement, how they cluster within schools and generate different estimates of the conditional (on SES) ethnic maths achievement gap. This section describes how the indices created through these techniques are analysed and what data are used to create them. A brief description of the index creation techniques is provided in section 7.4.1 and a more technical discussion of the methods is presented in appendix A.5.

The comparison of the methods' assumptions on observable and unobservable variables is based on both the theoretical assumptions of each method and the estimated correlations with observed variables. For continuous variables, the estimations of the correlations are based on Pearson's moment correlation; polyserial correlations are estimated for correlations between ordinal and continuous variables, polychoric correlations are estimated for each pair of ordinal variables, and correlations between nominal variables are estimated using Cramer's V. To examine how the continuous version of the variables and the composite indicators are distributed among ethnic groups, the effect size of the differences in SES between White and minority students are calculated. Using the within- and between-school and LA models in section 5.2.1, these effect sizes are also calculated for the within-school gaps in SES and the effect of an increase in the school and LA proportion of minority students on SES.

A set of model fit indicators measures the ability to predict maths achievement. The fitted models are single-level linear regression models with maths test scores as the dependent variable and a (set of) SES variable as the predictor. These measures include the coefficient of determination (R^2), which measures the proportion of variation in maths test scores that can be explained by variation in the SES variables. However, R^2 increases with the number of predictors, and one comparison of interest simultaneously includes the raw variables for education, occupation and income. For this reason, the adjusted R^2, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are also included for comparison. Besides, the size of the effect of single variables and indices on maths achievement are compared using standardised linear regression coefficients.

The intraclass correlations (ICCs) of each variable and indicator are estimated using three-level hierarchical models to examine clustering within schools and LAs. The ICC measures the degree of correlation in the SES measure between students within the same schools and LAs. The models are linear for the continuous variables (including the indices) and ordinal logit models for categorical variables (assuming the ordering implied by the continuous version of the variable)[4].

The conditional overall ethnic achievement gaps are estimated using single-level linear regression with cluster-robust standard errors (Arellano, 1987). The conditional within-school and school and LA contextual effects of ethnicity are estimated using three-level (students within schools within LAs) random intercept models. The reasons for choosing these estimation strategies are further studied in chapter 8, which explains why single-level models are preferred for studying the overall gap and multilevel mod-

[4]For more information about hierarchical models and the intraclass correlation coefficient, please refer to appendix 5.2.2.

els for each of the components of the gap. Both methods allow examining the effect of the SES's operationalisation on the estimated conditional ethnic achievement gaps.

This chapter uses data for the 2008 SABER 11 data cohort, which consists of 387,773 students in 7,143 schools and 80 LAs. The data cleaning procedures are the same as described in the data chapter. Additionally, since most of the index creation methods that are compared in this chapter require continuous data, the original SES variables in the SABER 11 database are transformed from their original categorical levels of measurement to continuous levels. There are many possible ways to perform this transformation, which may affect the outcomes that are analysed in this chapter. Nonetheless, understanding how these transformations affect the analysis is out of the scope of this chapter. Instead, the focus is on the methods that are used to create the composite indicators. For this reason, the variables are transformed from categorical to continuous in only one way. Implicitly, this assumes that all the index-creation methods are affected in the same way by the transformation of categorical into continuous variables.

As with the index creation methods, the way to transform categorical into continuous variables mirrors the literature review in section 7.2.2. This means that education, which is measured initially as categories of educational attainment, is transformed into years of education; occupational categories are transformed into occupational prestige using the Ganzeboom et al. (1992) index of occupational status; and family income is assigned the midpoint value of the original interval. Theses transformations only add meaning to the differences between categories. Ideally, when using methods that require continuous data, the variables should follow a continuous distribution. Nonetheless, these transformations are adopted in the interest of following the procedures that are conventionally followed in the literature.

There are two important differences when transforming occupational categories into occupational prestige. The first one is that, unlike other databases, which collect occupational data using a large number of occupational categories (based on census or international classifications), the SABER 11 database only collects occupation data in 12 broad categories. This implies that results using the SABER 11 database may be different from those using databases with richer occupational information. The second difference is that conventionally in the literature, pensioners and stay-at-home parents are not assigned a prestige score; instead, these categories are understood as missing data that are later recovered using imputation techniques. In contrast, in this chapter, pensioners and stay-at-home parents are assigned a prestige score of zero. This is because it allows using the same sample to construct all the indices to ensure their comparability. The drawback of this decision is that it creates outliers, which are likely to affect the composite indicators. Again, as long as all indicators are similarly affected by outliers, this does not affect the conclusions of this chapter.

As the results section includes tables and figures with abbreviations for the variable names, Table 7.2 presents the variable names along with their description. These include the original variables, which are used as a baseline for comparison, and the transformed variables. In general, continuous variables include 'cont.' in their name, and the (original) categorical variables include 'cat.'.

Table 7.2: Description of the variables included in the analysis

Variable	Description
father_ed cat.	Original (categorical) father's educational attainment
father_ed cont.	Father's years of education
mother_ed cat.	Original (categorical) mother's educational attainment
mother_ed cont.	Mother's years of education
father_oc cat.	Original (categorical) father's occupation
father_oc cont.	Father's occupational prestige
mother_oc cat.	Original (categorical) mothers' occupation
mother_oc cont.	Mother's occupational prestige
income cat.	Original (intervals of) family income
income cont.	Family income as midpoint of the original intervals
Sisben	Sisben classification level

Figure 7.2 shows the distribution of these transformed variables (which is the same as the distribution of the categorical version of the variables shown in section 4.3.1, but with numbers instead of categories). The figure shows that the distributions of mothers' and fathers' educational attainment are similar for all ethnic groups. Besides, mothers of all ethnicities are predominantly stay-at-home parents. In general, Afrocolombian and Indigenous students tend to have a more socioeconomically disadvantaged background than other minorities and White students[5]. Therefore, it would be expected that

[5]Table A.16, in appendix 8 shows the mean, standard deviation and median of the continuous version of the SES variables for all students and by ethnic group.

a composite measure of SES reflects these differences in SES between ethnic groups.

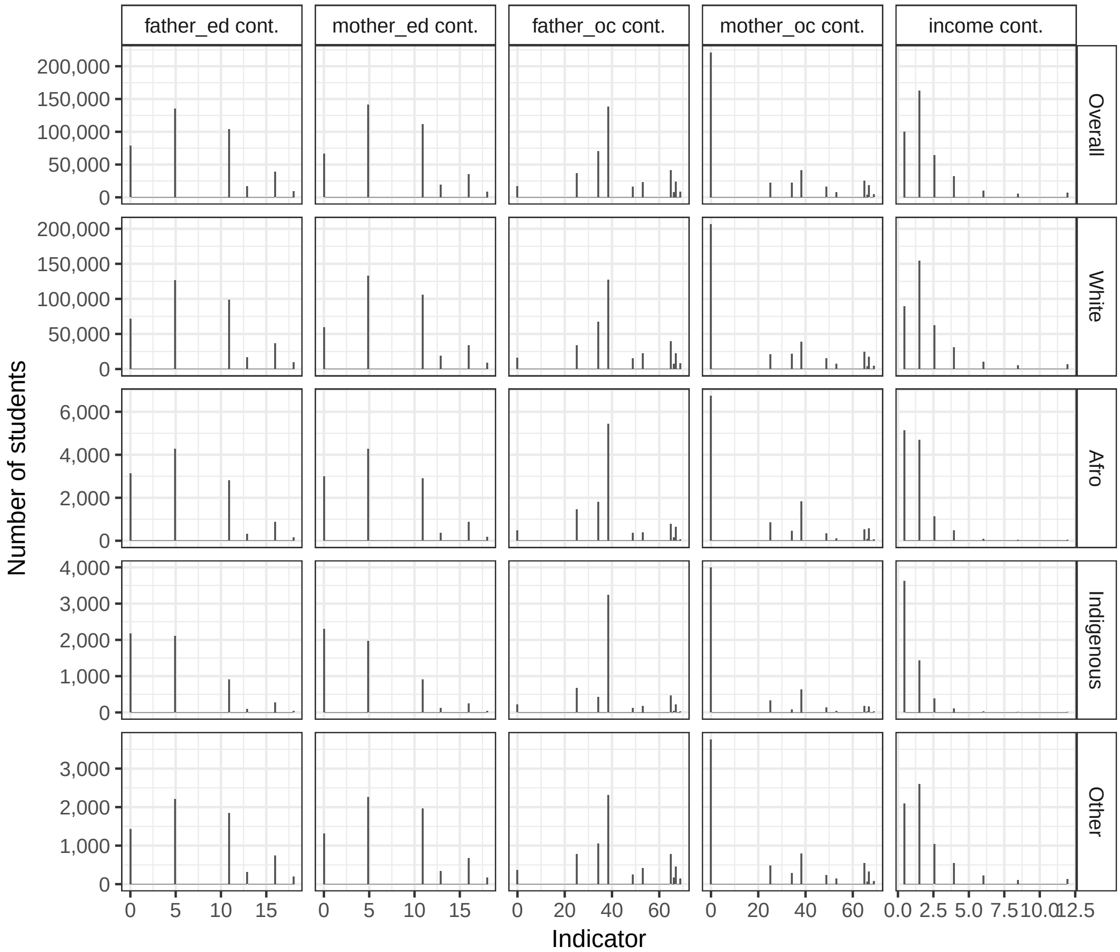

Figure 7.2: Distribution of the SES transformed by ethnic group and overall

7.4 Results

This section presents the results of comparing competing ways of combining education, occupation and income to measure SES using the SABER 11 database. As shown in section 7.2.2, there are broadly three ways of operationalising SES. First, including these three variables as controls in regression models directly. Second, combining them through some index creation technique to obtain a single SES indicator. Third, using a variable that is used to target public policy to disadvantaged students, such as FSM status. The equivalent to this variable in Colombia is the System for the Selection of Social Programs' Beneficiaries (Sisben) classification, as further described in appendix

A.1[6].

The following results compare these three alternatives, considering the index creation techniques that are commonly used in the literature and their implications regarding assumptions, distribution among ethnic groups, ability to predict maths achievement, impact on clustering estimation and the estimated conditional ethnic achievement gap. These index-construction techniques are:

- Factor analysis (FA)
- Principal component analysis (PCA)
- Average of standardised variables (s. average)
- Sum of standardised variables (s. sum)
- Sum of unstandardised (raw) variables (sum)
- Hollingshead (2011) four-factor index (Hollingshead)
- Prediction from linear regression (Prediction cat. for the index based on the original categorical variables and Prediction cont. for the index based on the transformed version of the variables).

7.4.1 Relationships Among Education, Occupation and Income

Following Cowan et al. (2012), the composite indicators in this chapter can be split into two groups: formative and reflective indicators. On one side, the formative indicators are those that assume that education, occupation and income blend together to create (cause) SES. This group includes the indicators created as the average or sum of raw or standardised variables, the Hollingshead (2011) four-factor index and prediction from linear regression. In all these cases, SES is understood as the combination of the observable variables education, occupation and income. The difference is how this combination is performed.

The sum of unstandardised variables method assigns, in theory, equal weights to each of the observed variables, which are measured on different scales (years of education, prestige and number of minimum wages). Therefore, the variable with the largest

[6]Please note that Sisben is used for comparison, but not used to create the composite indicators.

scale dominates the index[7]. In practice, this means that the weight of each variable depends on its scale. In turn, the sum and average of standardised variables assign equal weights to the standardised variables. The weight on the standardised variables is one when variables are summed and $\frac{1}{5}$ when the average is computed[8], which is equivalent to weighting each of the unstandardised variables by the inverse of their standard deviation, applying the weights (1 or $\frac{1}{5}$) and then re-centring the index. This is better shown in Table 7.3.

In turn, the Hollingshead (2011) four-factor index and the prediction from linear regression method assign different weights to each variable. The former assigns arbitrary weights of five to educational attainment and three to the occupational status[9], and is the only indicator that does not include income. The index created through linear prediction, in turn, assigns weights to each variable according to their coefficient in a linear regression of maths on parental education and occupation and family income. This is equivalent to predicting maths achievement according to these variables (minus the intercept). This method is the only one of the analysed in this chapter that requires maths test scores for its creation. The advantage of this method is that, unlike the methods above (requiring continuous variables), the observed variables can be measured either as categorical or as continuous variables.

On the other side, reflective indices understand SES as a latent variable that causes education, occupation and income to be correlated. PCA uses the observed variance between these variables to find components (in this case, the first component) that account for as much of the total variance as possible through a matrix-decomposition method. In turn, FA finds the factors that match the factors' correlations with the observed vari-

[7]For example, if education were measured in years (0 to 18 years) and income were measured in Colombian pesos (0 to \$10,000,000), income would dominate the final index. That is, a change from no education to postgraduate education (0 to 18 years) would have a smaller effect on the overall index than a marginal change in income (0 to \$100). In this chapter, income is measured as the number of minimum wages, between 0.5 and 12 minimum wages in the continuous version used to create these indices.

[8]This is because there are five variables in the index: father's education, mother's education, father's occupation, mother's occupation and income. The average is calculated as the sum of the values for each variable divided by the number of variables (five).

[9]The other two factors are sex and marital status. However, they are not directly used to compute the index, as explained by Hollingshead (2011). The index is calculated according to the member of the family that is employed outside the household. If both, mother and father are employed, the index for the family is the average of the mother's and father's indices.

ables' correlations as much as possible, using maximum likelihood estimation. If the correlation structure of education, occupation and income is strong enough, the indices created through these two methods should be very similar (Bartholomew, Steele, Galbraith, & Moustaki, 2008, p.200). Both methods require continuous variables. Although there are alternative reflective-index creation methods that can be used with categorical data, PCA and FA are more commonly used in the literature, as discussed in section 7.2.2.

Comparing these reflective methods with the only formative method that does not assign arbitrary weights to the variables (the prediction from linear regression method), there are two main differences. First, while the reflective indices assign weights to the observable SES variables in such a way they 'predict' the latent variable SES, the prediction from linear regression method assigns weights to the observable variables in such a way that they generate the best possible predictions of maths achievement. The second difference is that the prediction from linear regression method can be used with either continuous or categorical data.

Table 7.3 shows the weights that are assigned to each observable variable[10] using each method for creating composite indicators[11]. As shown in the table, FA, PCA and prediction from linear regression tend to assign higher weights to parental education (both, mother's and father's) than to any other variable. The least important variable is usually parental occupation. The first factor of FA explains 39.7% of the total variance, while the first principal component of PCA explains 50.3% of the total variance of the input variables[12].

These weights lead to the distributions of the indices that are depicted in Figure 7.3. As shown in the figure, the indices differ in skewness with the FA index being more symmetric than the indices created by prediction. Nonetheless, all indicators show differences in the distribution of SES among ethnic groups, with Indigenous students tending to have lower SES. This result is better visualised in Figure 7.4, which shows

[10]Please note that the table presents the equivalent weights for the unstandardised version of the variables, while FA and PCA were carried out using standardised variables to avoid the loadings to be affected by the scale of the variables.

[11]The Hollingshead (2011) four-factor index and a second version of the index created using prediction from linear regression have been omitted from the table as they are based on categorical versions of the data. The last ones are shown in appendix A.5.

[12]The standard output for FA and PCA can be found in appendix A.5.

Table 7.3: Weights for observed variables by index

	father_ed cont.	mother_ed cont.	father_oc cont.	mother_oc cont.	income cont.
FA	0.148	0.159	0.023	0.018	0.299
PCA	0.096	0.101	0.021	0.015	0.231
s. average	0.037	0.039	0.012	0.008	0.101
s. sum	0.187	0.195	0.062	0.040	0.503
sum	1.000	1.000	1.000	1.000	1.000
prediction cont.	0.013	0.018	0.002	0.001	0.099

the size of the gap between White and minority students in SES according to the individual transformed variables and composite indicators, measured as the standardised coefficient of a regression of SES on ethnicity. As with the overall ethnic achievement gap examined in chapter 6, the ethnic SES gap can be decomposed into its within-school and between-school and LA components, which are also shown in Figure 7.4. Here, the focus is on the constituent parts (the within-school SES gap and the school and LA contextual effects of ethnicity on SES) and the measurement of SES by regressing SES on ethnicity and the school and LA proportions of minority students.

Consistently with the description in chapter 4, Figure 7.4 shows that Indigenous students tend to come from more disadvantaged backgrounds, in terms of both parents' education, mother's education and income. This is picked up by all the indicators, although the size of the difference in average SES between White and minority students (the ethnic SES gap) varies between 0.02 standard deviations (SD), when SES is operationalised as mother's occupational prestige, and 0.08 SD when using the prediction from linear regression method for categorical variables. However, while all indicators point out that Afrocolombian and Indigenous students come from more disadvantaged backgrounds than White students, other minority students have fathers with more years of education, mothers in more prestigious occupations, and higher SES according the sum of unstandardised variables and the Hollingshead (2011) four-factor index, in comparison to White students.

In general, the examined operationalisations of SES show that differences in SES between White and minority students are much smaller (close to zero, according to all indicators) within schools than between schools and LAs. Nonetheless, the choice of indicator matters when estimating the magnitude of the school and LA contextual effects of ethnicity on SES. For Afrocolombian students, parental education estimates the

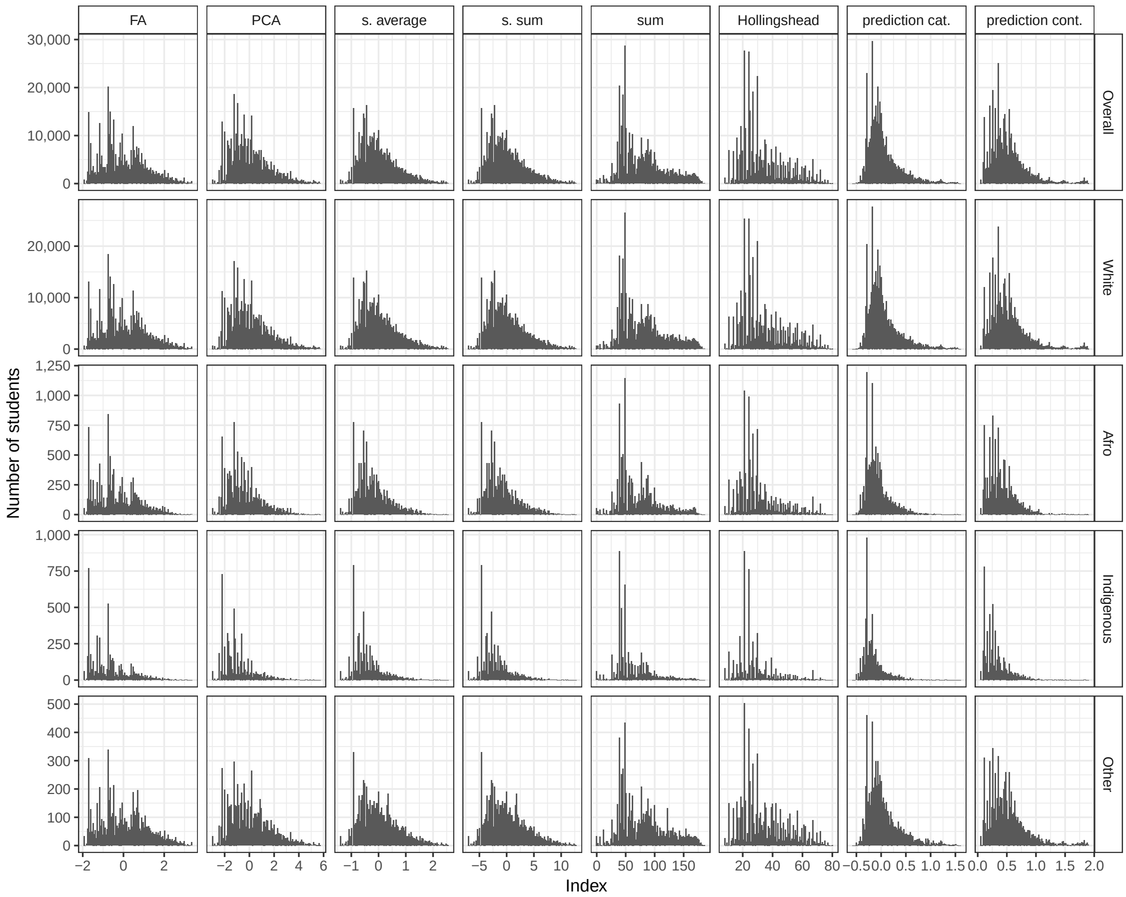

Figure 7.3: Distribution of the SES composite indicators overall and for each ethnic group

smallest school contextual effect (0.08 SD) while the index created by prediction from linear regression estimates the largest one (0.1 SD). In turn, for Indigenous students, the Hollingshead (2011) four-factor index and the raw sum of variables estimate the smallest school contextual effect (0.1 SD) and FA and similar indicators created using standardised variables estimate the largest one (0.13 SD). Parental education and the index created by prediction from linear regression result in the smallest and largest LA contextual effect on SES for both Afrocolombian (0.08 SD and 0.16 SD, respectively) and Indigenous students (0.28 SD and 0.37 SD, respectively). For other minority students, the SES gap and its components are very close to zero, regardless of the operationalisation of SES. Therefore the importance of the choice of SES when calculating differences in SES between ethnic groups varies both, by ethnic group and by the level of the education system under analysis.

To better understand how these SES indicators correlate with the observable vari-

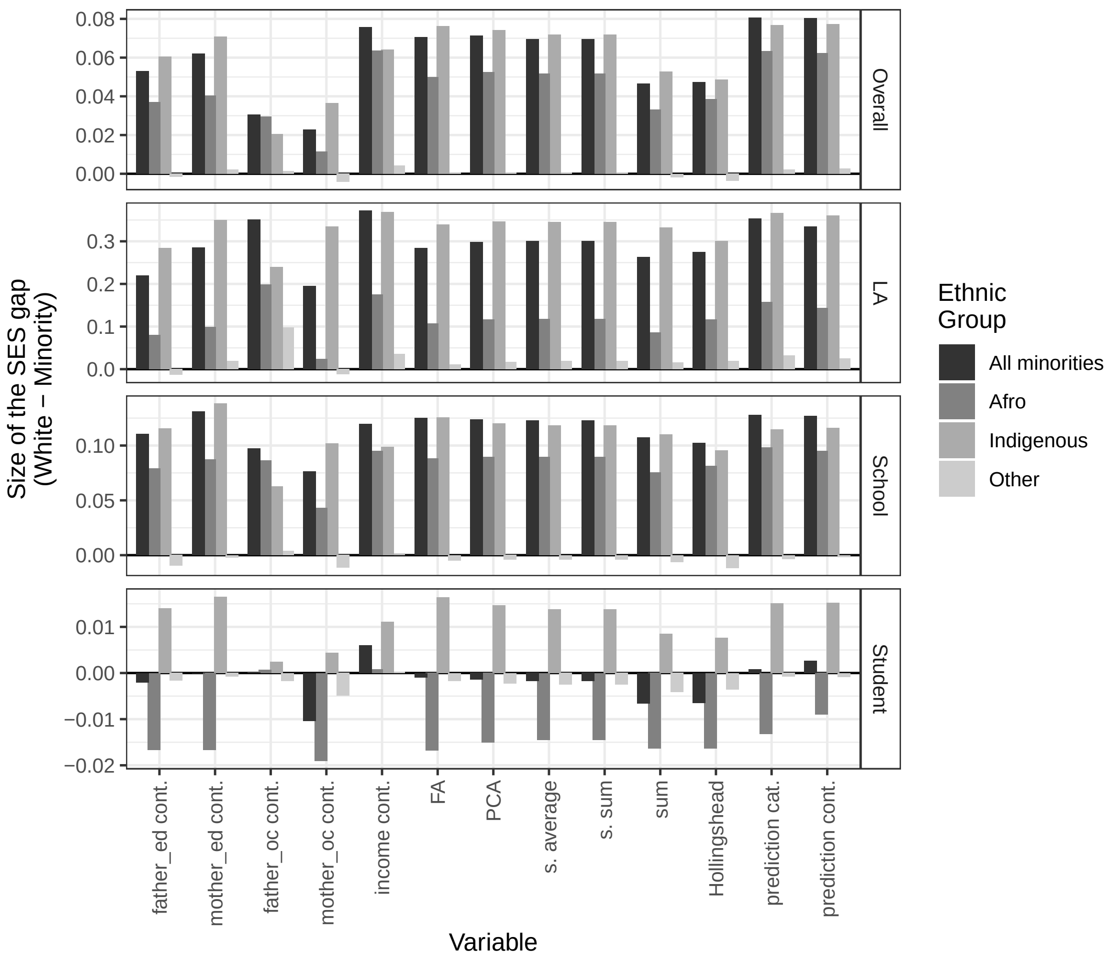

Figure 7.4: Size of the differences in the overall, within-school (student) and between school and LA mean SES by ethnic group and operationalisation

ables and each other, Figure 7.5 presents the correlation matrix between observed variables and SES indicators. The indicators, including Sisben tend to have a stronger correlation with education (however measured) than with occupation and income, and the correlation with occupation tends to be the weakest one, reflecting the weighting scheme for the indices. Two exceptions, as expected, are the Hollingshead (2011) four-factor index and the unstandardised sum of variables, which show a higher correlation with parental occupation[13]. Besides, all composite indicators are very highly correlated, in particular, FA, PCA and the sum and average of standardised variables. When examining these correlations by ethnic group, as shown in appendix A.5, the results are very

[13]In the case of the unstandardised sum of variables, a higher correlation with occupation was expected as this variable takes values between 0 and 69, whereas education varies between 0 and 18 years of education, and income between 0.5 and 12.

similar.

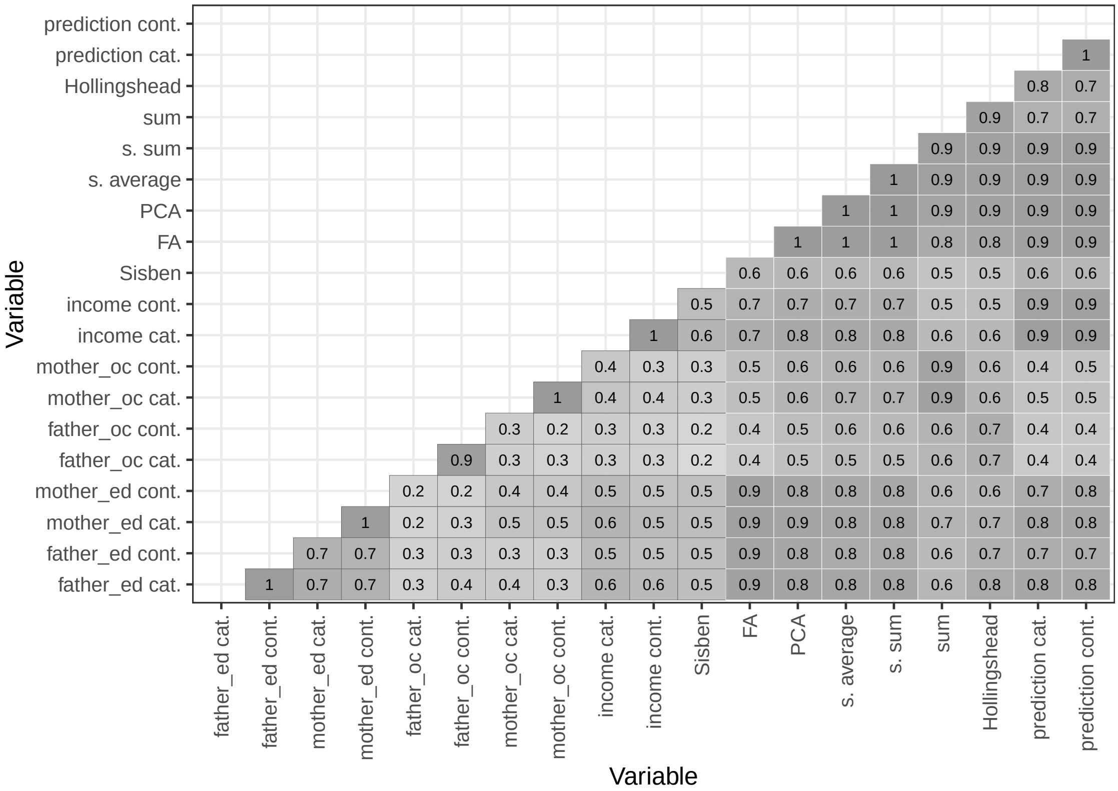

Figure 7.5: Correlations (to the nearest 0.1) between observed variables and SES composite indicators

Given these results, it would be expected that the indicators perform very similarly as predictors of achievement and when estimating the ethnic achievement gap conditional on SES. The next sections examine whether this is the case.

7.4.2 Differences in Predictive Power

Three model fit indicators are used to check what variables are better predictors of maths achievement: the adjusted R^2 (adj.-R^2), the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The higher the values of the adjusted R^2 and the lower values of the AIC and the BIC, the better the model fit. These indicators of model fit are only comparable when estimated on the same sample, which means that comparisons like 'income is a better predictor of achievement for White students than for minorities' are not possible, since the sample of White students, for example, is different to the sample of Afrocolombian students. What is possible is to explore if the best predictor of maths achievement for White students is the same for minority students.

Figure 7.6 shows that this is indeed the case; all model fit measures indicate that prediction from a linear regression using categorical variables provides the best model fit, with an adjusted R^2 of 0.1 for the whole sample. This is because this index explains as much variance of maths test scores as including all the original categorical variables in the model, but only requires one degree of freedom.

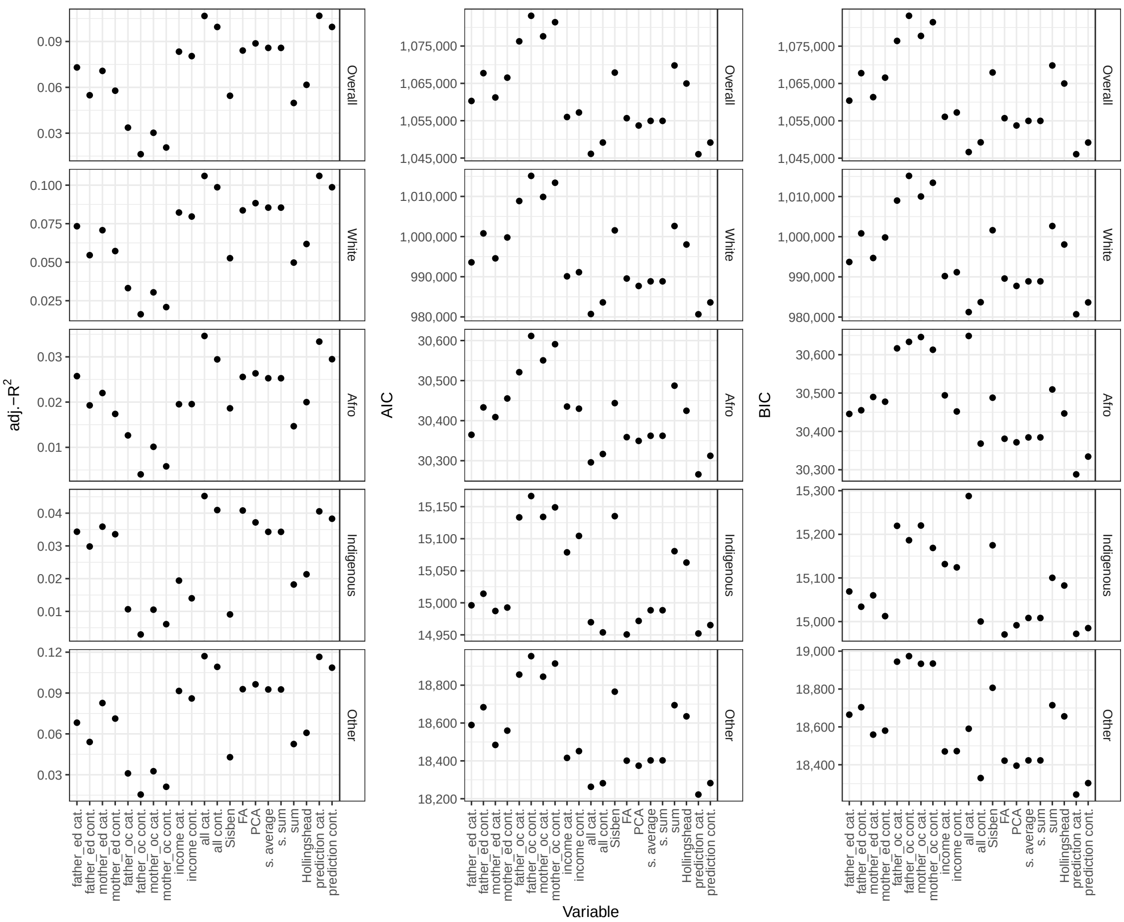

Figure 7.6: Model fit for observed variables and SES composite indicators overall and for each ethnic group

In turn, including all categorical variables in the model provides the next best model fit (with a marginal difference in the adjusted R^2), but at the cost of 46 degrees of freedom, given the total number of categories that need to be included in the model to account for both parents' education and occupation, and family income. This is heavily penalised by the BIC in the minority groups with the smallest sample sizes

(Afrocolombian, with a BIC of 30,649 and Indigenous students, with a BIC of 15,288).

Regarding individual variables, income is the strongest predictor of maths achievement for White (adj.-R^2 =0.08) and other minority students (adj.-R^2 =0.09), while it is parental (father's and mother's) education for Afrocolombian (adj.-R^2 =0.03) and Indigenous students (adj.-R^2 =0.04). The figure also shows that, in general, the original categorical variables offer a better model fit than the transformed version of the variables. Overall, Sisben is not the best predictor of maths achievement (adj.-R^2 = 0.02, 0.01 and 0.04 for Afrocolombian, Indigenous and other minority students, respectively). For all ethnic groups, another single-variable predictor would be a better alternative, as discussed above.

In terms of other composite indicators, unsurprisingly FA, PCA and the sum and average of standardised variables have very similar model fit (with an overall adjusted R^2 around 0.08). The raw sum of variables and the Hollingshead (2011) four-factor index have a poorer model fit, comparable to using a single variable to measure SES (overall adjusted R^2 around 0.05).

An alternative that allows comparing the strength of the correlation between maths achievement and the SES among ethnic groups is to compute the correlation coefficients between these variables. The disadvantage in this case is that the approach of using all observed SES variables cannot be compared. Figure 7.7 shows that, indeed, the strength of the linear relationship between SES indicators and maths achievement varies by ethnic group although, in general, SES is not as strongly correlated with maths achievement for Afrocolombian (with correlations between 0.05 and 0.18 for parental occupation and the index created by prediction of linear regression, respectively) and Indigenous students (with correlations between 0.04 and 0.2 for parental occupation and the index created by FA, respectively) as it is for White and other minority students (with correlations between 0.11 and 0.12 for parental occupation, and 0.33 and 0.34 for the index created by prediction from linear regression, respectively). Nonetheless, indicators such as income generate larger differences in this correlation between Afrocolombian and Indigenous and White and other minority students, with correlations of 0.14 and 0.12, and 0.28 and 0.29, respectively.

The ability to predict maths test scores is one of the properties that has been used to choose between potential measures of SES (Braveman et al., 2001; Ilie et al., 2017). However, using this as the only criterion for choosing a measurement of SES may hide

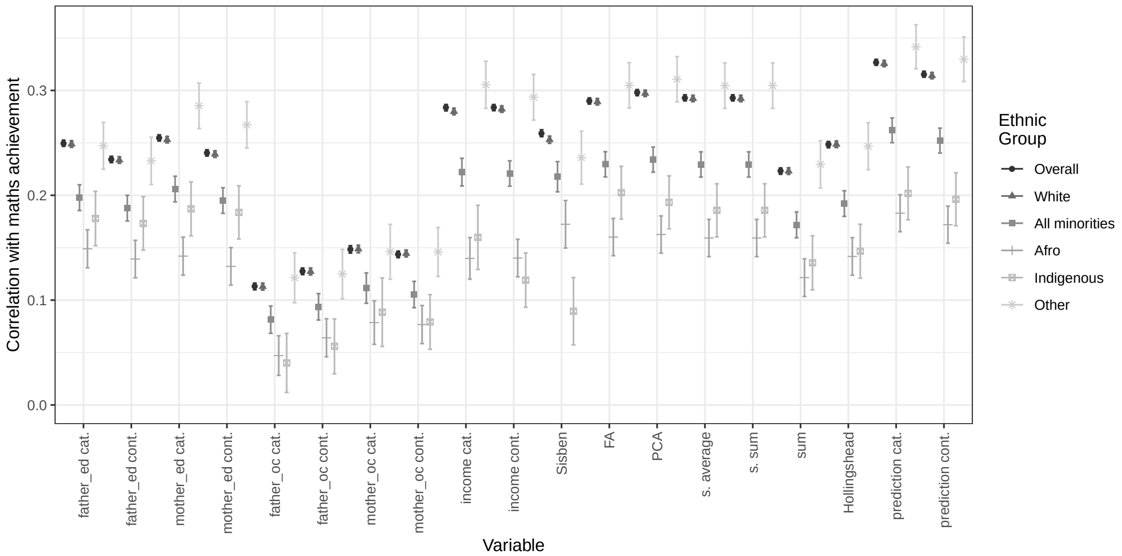

Figure 7.7: Correlations between SES variables and maths achievement overall and for each ethnic group

differences in other areas under study. The next sections review two of these areas: the clustering within schools and LAs and the ethnic achievement gap.

7.4.3 Differences in Clustering Patterns

In order to understand whether different SES measures produce different estimates of the clustering of students within schools and LAs according to their SES, this section examines the intraclass correlation (ICC) of the different SES measures, including individual variables. Clustering of SES on its own can be considered a measure of inequality or school and LA segregation. Besides, the clustering of SES may change the amount of unexplained variability in maths test scores, affecting fixed and random variability in the gap, as further explored in chapter 9. The last is the reason to consider the clustering of the different indicators in this chapter.

Overall, the indicators suggest a high level of within-school clustering. Figure 7.8 shows that very different estimates of the clustering of SES can be obtained depending on how SES is measured. The SES intra-school correlation estimated by the categorical variables of education and income (around 0.7 for all of them) tend to be higher than the correlation estimated by the transformed continuous variables (0.6 for both parents' education and 0.4 for income), but there is no much difference between the categorical and continuous versions of occupation (around 0.8). These correlations are very similar for all ethnic groups.

Figure 7.8 also shows that the correlation in SES between schools within the same LA is generally low (estimated around 0.3 and 0.4 for the categorical and continuous versions of parental education), with the exception of the continuous version of income (with an estimated correlation around 0.6, twice the estimated correlation for the categorical version of the variable). This difference is likely to reflect the choice of the method to transform the income variable from its original categorical version to the continuous version. The correlation in SES between LAs is much smaller, estimated below 0.1 for parental education and occupation and family income, with the same results for all ethnic groups.

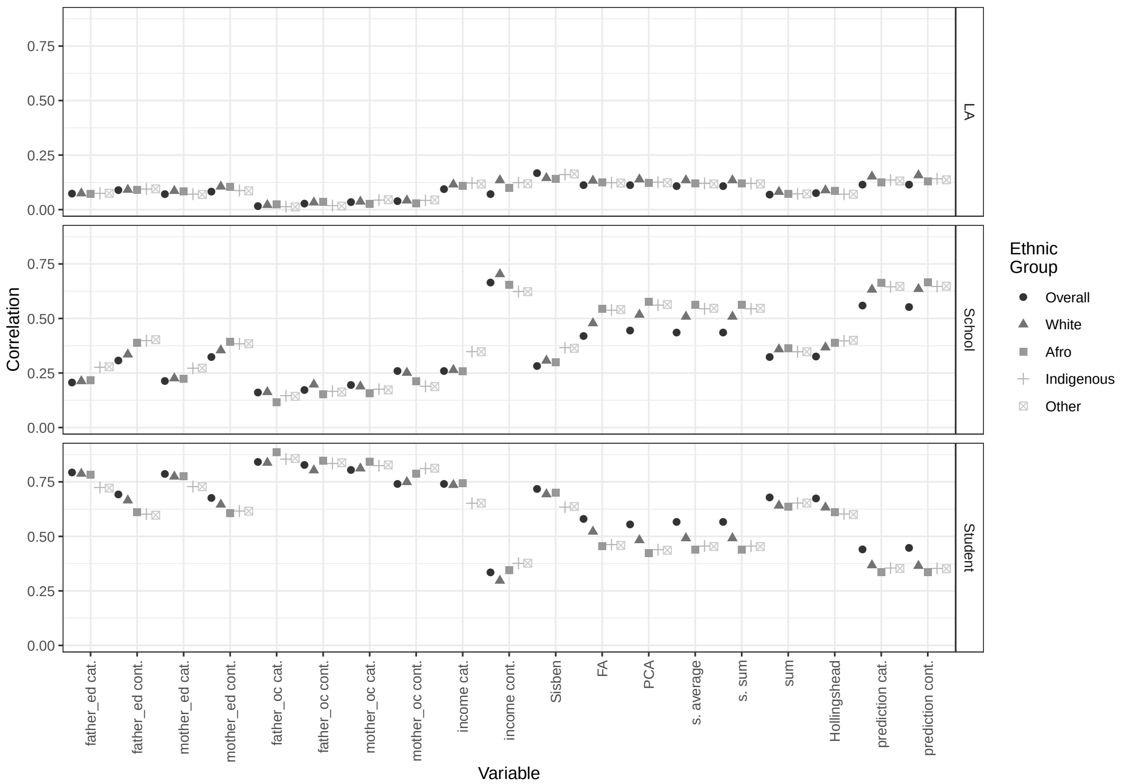

Figure 7.8: Intra-group correlations in observed variables and SES composite indicators for students, schools and LAs overall and for each ethnic group

As shown in Figure 7.8, all the composite indicators, excepting those created by linear prediction, reflect the within-school clustering of the observed variables, as theirintra-school SES correlations are bounded by the minimum and maximum correlation of their components, varying between 0.4 and 0.7. However, those created using FA, PCA and the sum and average of standardised variables report a very similar intra-school SES correlation (0.5), which is lower than those created using the sum of variables and the Hollingshead (2011) four-factor index (0.7).

In this case, the aggregation of the variables to create composite indicators tends to favour between-school as opposed to within-school variability, and therefore the indices tend to generate estimates of lower within-school clustering and higher SES correlations between students in different schools and LAs. The effect of aggregation seems to be lower for unstandardised variables, and the same results hold for all ethnic groups. For example, the index created by the standardised sum of variables leads to an estimate of a 0.5 intra-school SES correlation, which is lower than the 0.7 estimate of the correlation of an index created by the sum of unstandardised variables[14], which is more similar to the within-school clustering of the raw variables. Finally, the indicators created by linear prediction lead to the lowest estimates of SES within-school clustering (0.4). This is likely to be the case because, as they are created according to the maths test scores, they may pick up some of the between-school and -LA variation in test scores and incorporate it into the indicator.

These changes in the within-school and between-school and LA clustering patterns are expected to affect the estimation of random-slope models for the ethnic achievement gap studied in chapter 9. Nonetheless, these patterns, along with patterns in the SES gap between White and minority students (studied in section 7.4.1) and the differences in the ability of the indicators to predict maths achievement (in section 7.4.2) may influence the estimates of the conditional ethnic achievement gaps and their components, as studied in the next section.

7.4.4 Differences in the Conditional Ethnic Achievement Gaps

This section presents the effects of different SES operationalisations on the estimated conditional ethnic achievement gaps and their components (the within-school gaps and school and LA contextual effects of ethnicity). Since minority students have a lower average SES (as shown in section 7.4.1) and SES is positively correlated with maths achievement (as shown in section 7.4.2), controlling for SES should narrow the estimated overall ethnic gap. The results presented here focus on the changes in the gap, while appendix A.5.2 shows the estimated conditional gaps. How this effect splits within- and

[14]Please notice that it is the effect of the aggregation of standardised variables and not of the standardisation of the variables, as the ICCs of standardised and unstandardised variables are the same.

between-schools and LAs is determined by how the students' ethnicity relates to their SES, how the schools' and LAs' ethnic composition relates to their SES composition and how all these relate to maths achievement.

As estimated in chapter 6, the unconditional achievement gap between White and minority students is 0.35 SD, while the within-school gap is 0.06 SD and the school and LA contextual effects of ethnicity (the proportion of -pooled- minority students) are 0.22 SD and 0.41 SD, respectively. Figure 7.9 shows the percentage change in these estimates (for all minorities together and by ethnicity) after controlling for SES using different variables and composite indicators.

The estimated maths achievement gap is narrower when comparing White and minority students with the same SES. Reflecting the patterns found in the previous sections, the operationalisation of SES affects the proportion by which the gap narrows. If SES is operationalised as a single variable, the gap decreases between 3.9%, when using parental occupation and 29.3% when controlling for Sisben (to 0.33 SD and 0.25 SD, respectively). In contrast, when SES is operationalised by jointly including individual variables for parental education and occupation and family income in the regression model, the achievement gap narrows by 30% (to 0.24 SD). This decrease is almost the same than the one estimated when operationalising SES as the prediction from linear regression (31%, to 0.24 SD), but higher than the one resulting from operationalising SES as the sum of unstandardised variables or the Hollingshead (2011) four-factor index (13.9%, to 0.3 SD) or as one of the indicators created with standardised variables (24%, to 0.26 SD).

As in chapter 6, the overall ethnic achievement gap hides additional information about the patterns of ethnic inequality. Considering different levels of the education system (students, schools and LAs) reveals complex relationships between SES indicators, ethnicity and maths achievement, as shown in Figure 7.9. If well controlling for SES using composite indicators leads to a larger change in the estimated achievement gap, these changes are different for each level. For example, controlling for SES using only parental education leads to a 6% increase in the within-school gap between White and minority students (to 0.06 SD), a 11.6% decrease in the school contextual effect of minority students (to 0.2 SD), and a 5.7% decrease in the respective LA contextual effect (to 0.38 SD).

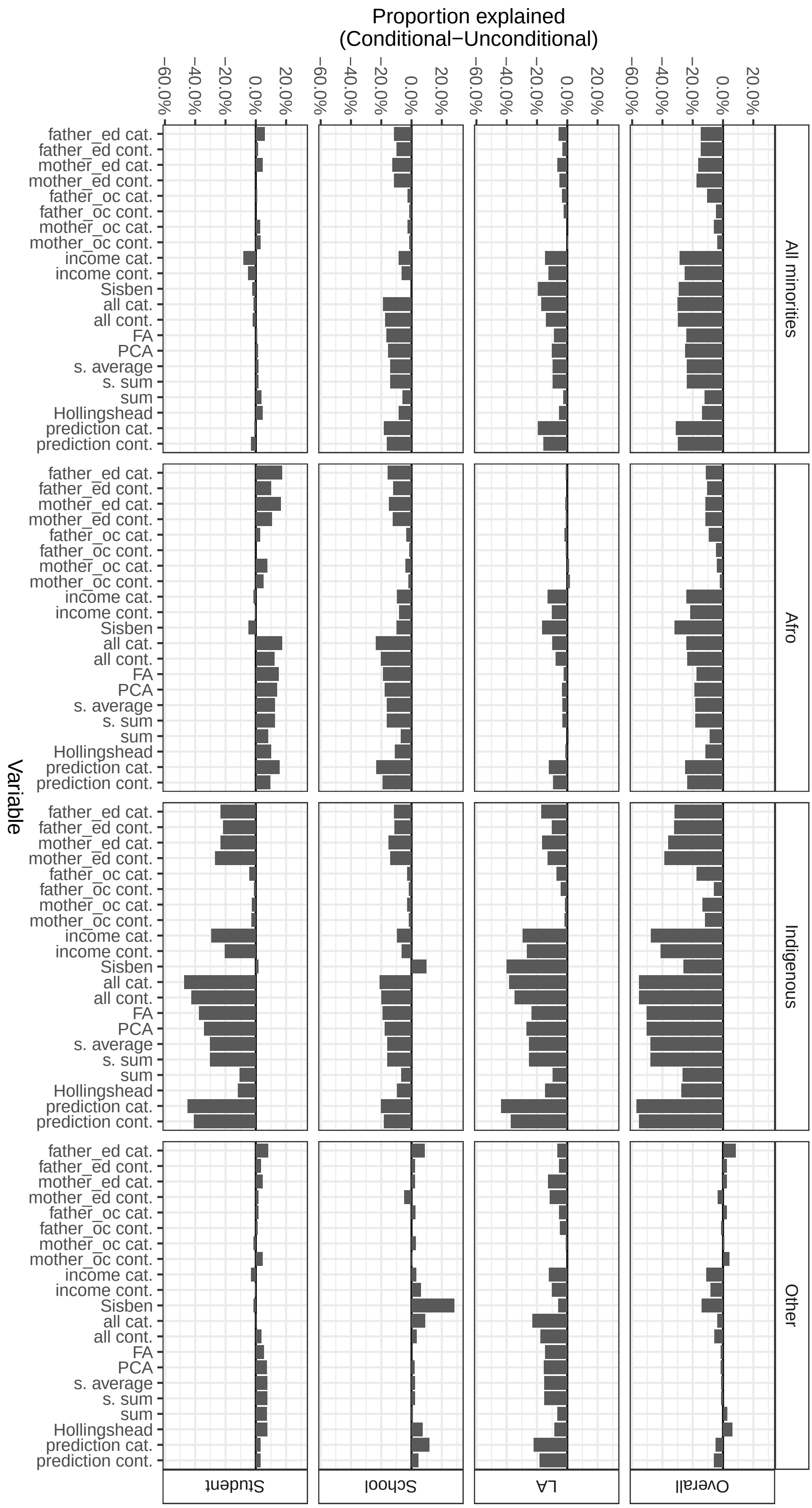

Figure 7.9: Percentage change in the estimated ethnic maths achievement gaps and their components when controlling for SES by indicator

In contrast, controlling for SES using prediction from linear regression leads to a 1.1%, 18.4% and 19.3% decreases in the within-school gap and school and LA contextual effect of ethnicity (to 0.06 SD, 0.18 SD and 0.33 SD), respectively. Nonetheless, the results are similar within the groups of indicators that have been identified (i.e. those that use standardised variables and those that use raw variables).

Figure 7.9 also shows that differences among ethnic minorities are important when analysing the effect of choosing how to operationalise SES. Even though the indicators consistently show that SES is more important to explain the achievement gap between White and Indigenous students than the gap for Afrocolombian and other minority students, the choice of SES indicator seems to be more important for the White-Indigenous gap than for any other gap. In the case of Afrocolombian students[15], even though individually controlling for parental education or occupation explains between 4.5% and 11.4% of the gap (leading to a conditional gap between 0.43 SD and 0.46 SD), controlling only for family income or simultaneously controlling for parental education and occupation and family income generate similar estimates of the conditional ethnic achievement gap (0.36 SD), accounting for 24.1% of the overall achievement gap. In other words, family income seems an appropriate single indicator to control for SES when studying the achievement gap between Afrocolombian and White students. There is also a small difference between using income, including all observed variables directly in the regression or the composite indicators created using FA, PCA, and the sum and average of standardised variables (the last group produces an estimated gap of 0.4 SD, explaining 17.4% of the gap). Finally, the Sisben classification is the indicator that explains most of the gap between Afrocolombian and White students, as the estimated gap decreases 31.9%, to 0.33 SD.

The analysis of the components of the gap in Figure 7.9 reveals that this is the case mainly because Sisben is one of the few indicators that estimate a narrower conditional within-school achievement gap for Afrocolombian students (with a 4.8% decrease, to 0.09 SD), and explains a larger proportion of the LA contextual effect of Afrocolombian students (16.5%, with an estimated conditional LA contextual effect of 0.25 SD) than alternative operationalisations (which explain between 2.4% and 9.9% of the contextual

[15] Afrocolombian students score on average 0.48 SD lower than their White peers, and 0.1 SD than White students in the same school. The school and LA contextual effects of Afrocolombian students are 0.29 SD and 0.3 SD, respectively.

effect, estimating it to be between 0.27 SD and 0.29 SD). Therefore, although the choice of SES operationalisation does not have a major effect on the estimation of the conditional overall gap, this choice affects the analysis of the components of the gap.

When estimating the conditional maths achievement gap between Indigenous and White students[16], Figure 7.9 shows that using parental education or family income to measure SES produces similar estimates of the gap (0.23 SD for mother's education and 0.19 SD for family income, explaining 36.2% and 47.6% of the overall gap, respectively). In turn, parental occupation produces an estimated conditional ethnic gap of around 0.3 SD, explaining 17.5% of the overall gap. However, the estimated conditional gap simultaneously using parental education and occupation and family income to measure SES (0.16 SD) is different from these estimates, as it explains 55.4% of the overall gap.

This implies that, unlike what happens with the conditional gap between Afrocolombian and White students, using a single variable to control for SES when studying the overall conditional gap between Indigenous and White students does not produce similar results to those when measuring SES with all available observed variables. Regarding the use of composite indicators, the results are the same as to the ones of the Afrocolombian achievement gap; the indicators created by FA, PCA and the average and sum of standardised variables produce similar estimates of the conditional gap (0.18 SD, explaining 50.4% of the gap) as those produced by controlling for the set of observed variables, which differ from the estimates when using the sum of unstandardised variables or the Hollingshead (2011) four-factor index (0.26 SD), as Figure 7.9 illustrates. When examining the components of the gap for Indigenous students, the patterns are similar to those found for Afrocolombian students, with the indicators created using standardised variables explaining a higher proportion of the school contextual effect than other possible operationalisations, and the inclusion of individual or a set of raw variables explaining a higher proportion of the LA contextual effect than competing indicators.

As shown in Figure 7.9, controlling for SES does not have a big impact in the gap between White and other minority students. This is expectable as section 7.4.1 showed that these students come from a similar or more advantaged SES background than White students. However, the variable that is used to control for SES can potentially change the conclusions of a study focusing on these studies. In particular, when controlling for

[16]The unconditional achievement gap between White and Indigenous students is 0.36 SD, the within-school gap is 0.05 SD and the contextual effect is 0.24 SD and 0.34 SD.

SES using the Sisben classification, the achievement gap reduces in 14.2%, to 0.09 SD. In turn, using the father's education would generate a conditional achievement gap of 0.12 SD (an increase of 8.4% in the unconditional overall gap.

Overall, these results show that the choice of SES operationalisation can potentially affect the results of the analysis, especially when examining the components of the gap (within-school gap and school and LA contextual effects) and not only the overall gap. Nonetheless, the transformation of the original categorical variables into continuous variables leads to similar results, given the approach to the conversion adopted in this chapter. The composite indicators created by linear prediction produce similar estimates to those of including all observed SES variables into the regression model. One limitation of this analysis is that it is not possible to test whether these differences are due to random chance or if they are statistically significant. Figure A.26, in appendix A.5 shows the confidence intervals for the percentage change in the achievement based on Tofighi, MacKinnon, and Yoon (2009) approximation. The figure is not part of the main analysis because this approximation led to the estimation of negative variances, and therefore, its applicability to this research context is questioned. Future research can explore the option of using bootstrapping to estimate these standard errors.

7.5 Discussion

7.5.1 Summary of Findings

The systematic literature review in section 7.2.2 showed that papers studying achievement gaps indexed in Scopus control for SES either using individual variables for education, occupation or income (or a set of them), a variable that is used for policy purposes (such as FSM in the UK or Sisben in Colombia), or composite indicators created using the following methods:

- Factor analysis (FA)
- Principal component analysis (PCA)
- Average of standardised variables
- Sum of standardised variables
- Sum of unstandardised variables

- Hollingshead (2011) four-factor index
- Prediction from linear regression

Unlike other reviews, the focus was on the methods used to create composite indicators, rather than on determining the size of the correlation between achievement and SES (Sirin, 2005; White, 1982) or on listing the variables that are used to measure SES (Buchmann, 2002; Ensminger & Fothergill, 2003). Nonetheless, this review is consistent with the existent reviews in that it shows the extensive support in the understanding that parental education and occupation and family income represent SES and that these variables are correlated with academic achievement (APA Task Force on Socioeconomic Status, 2007; Buchmann, 2002; Ensminger & Fothergill, 2003; Sirin, 2005; White, 1982).

The results in section 7.4 showed that, although most of the SES indicators lead to the same overarching conclusion (SES partially explains the ethnic achievement gap), the choice of indicator matters when estimating the magnitude of the importance of SES. The methods to operationalise SES can be grouped into five categories: First, including individual variables. Second, using a measure for public policy (Sisben in Colombia, which is comparable to FSM status in the UK). Third, the inclusion of all variables into the regression model and an index created by prediction from linear regression. Fourth, indices created using standardised variables (those created by FA, PCA and the average and sum of standardised variables), which give more weight to parental education. Finally, indices that do not use standardised variables (the Hollingshead (2011) four-factor index and the sum of unstandardised variables), which are loaded towards family income. These groups of indicators combine both, formative and reflective indicators, which means that despite the theoretical arguments for choosing one or the other (Caro & Cortés, 2012; Cowan et al., 2012), this is not a distinction that makes a difference under the context of this study.

This grouping was expected given that section 7.4.1 showed high correlations between SES operationalisations within these groups. In the measurement literature, such strong correlation is assumed to indicate that the variables are measuring the same underlying construct (validity) (Chapman et al., 2016). For individual variables, which have been compared in the literature before, section 7.4.1 showed that the correlation between education and income is around 0.6 and between occupation and income is around 0.4. These correlations are higher than those reported for countries in the Organisation for Economic Co-operation and Development (OECD) using the PISA 2000 data, which uses possessions as a proxy of wealth (0.2 for education and wealth and 0.3 for occupation and wealth) (G. N. Marks, 2011), but are similar to the correlations between education and income (0.6) and between occupation and income (0.4) reported using survey data for California and Cambridge, MA, US (Braveman et al., 2001; Kahl & Davis, 1955). Braveman et al. (2001) additionally showed that, unlike in Colombia, in California these correlations vary noticeably among ethnic groups, being as low as 0.3 for Latin and as high as 0.7 for American Indian groups.

In the case of OECD countries, the low correlations between education, occupation and income imply that the results of a study operationalising SES as a composite indicator may depend more on the weights that such indicator assigns to each of these variables than they are in the Colombian context. Additionally, in places where the correlations between SES variables vary among ethnic groups, such as California, it could be expected that the choice of SES indicators has a stronger effect on the conclusions of the analysis of ethnic differences. This chapter showed that, in Colombia, the overall conclusions about the conditional ethnic achievement gaps hold regardless of the SES operationalisation. Nonetheless, despite the similarity of correlations between SES variables among ethnic groups, it was not possible to identify a unique indicator that affected the estimation of the gap in the same way for all ethnic minorities. This is probably the case because, as shown in this chapter, the correlation between variables are not the only aspect that influences the effect of the operationalisation of SES.

Another of these aspects is the ability of these indicators to predict academic achievement, studied in section 7.4.2. This section showed that, consistently to White (1982)'s and Sirin (2005)'s findings, sizes of the correlations between maths and SES for all students varied between 0.11 and 0.33 according to the SES operationalisation. Nonetheless, the correlation between income and maths achievement in Colombia (0.28) is not a high as the correlation between home resources and achievement reported by White (1982) (0.57) and Sirin (2005) (0.51). G. N. Marks (2011) showed that for OECD countries participating in PISA 2000, the correlation between reading achievement and either parental education or occupation (0.33) was lower than a composite indicator combining parental education, occupation and home possessions using sheaf coefficients (0.43) and therefore recommended using composite indicators. The results in this chapter show, however, that this recommendation is dependent on the method to create such composite indicators, as those created using standardised variables estimate a correlation of 0.29, while these correlations are of 0.22 for the Hollingshead (2011) index and the sum of raw variables, and 0.33 for indices created using prediction from linear regression.

The results regarding the use of the Sisben classification instead of all SES variables contrast what was found by Ilie et al. (2017) in the case of the FSM status in the UK. While in the UK using only FSM to control for SES instead of using the whole set of variables explored by Ilie et al. (2017) implies 0.3% less of explanatory power at the student level, in the Colombian case, using the Sisben classification implies a reduction of 49% in the explanatory power of SES (measured by the R^2). Therefore, unlike Ilie et al. (2017)'s recommendation, it would not be advisable to use the Sisben classification to replace all the measures of SES.

Section 7.4.3 then estimated the clustering patterns of the different SES indicators within and between schools and LAs, showing that the SES; within-school clustering in Colombia is high (between 0.35 and 0.85, as estimated by an index created using prediction from linear regression and parental education, respectively), especially when compared to reports from Sweden[17], where the ICC for SES variables ranges between 0.03 and 0.18 (Yang Hansen, Rosén, & Gustafsson, 2011, p.201). Consequently, estimates of within-LA/between-school clustering are also affected by the operationalisation of SES.

At this point, it is difficult to assess which index-creation method provides a better estimate of the SES clustering, as it is an unknown parameter. However, it has been argued that ignoring the multilevel structure of the covariance matrix would lead to incorrect estimates of the factor loadings in the case of FA and it would be expected for it to lead to incorrect estimates of the clustering patterns (Goldstein & Browne, 2005; Hox, 1993; Longford & Muthén, 1992). The index created by prediction from linear regression may also result in smaller estimates of clustering because it is the only indicator that incorporates information about maths test scores, and single-variable operationalisations may omit important information from other variables that are also part of SES. Therefore, there is no clear indication of which intra-group correlations are 'correct', but this finding implies that the choice of SES indicator can potentially affect

[17]If well the Swedish and Colombian contexts are very different, this is comparison can be used as a reference of how unequal the Colombian education system is. Besides, authors do not usually report the ICC of SES. Normally only the ICC of the dependent variable is presented.

the analysis. In this case, those indicators with lower intra-school clustering may lead to the estimation of models indicating larger random between-school and LA variation in intercepts and slopes.

The differences found in the distribution of SES, predictive power and within-school clustering among SES indicators lead to different estimates of the conditional ethnic achievement gap, as studied in section 7.4.4. Therefore, the earlier distinction between five groups of indicators is also reflected in the proportion of the achievement gap that is explained by differences in SES. For individual variables, the overall gap between White and all (pooled) minority students decreases between 3.9% (operationalised as mother's occupation) and 28.8% (operationalised as family income); while these figures are 29.3% for Sisben, 31% for including the 'big three' simultaneously or using an index created by prediction from linear regression, 24% for the indices created by standardised variables, and 13.9% for the Hollingshead (2011) or the sum of raw variables, respectively[18].

The mechanisms through which each of these operationalisations combine differences in their distribution among ethnic groups, predictive power for maths achievement and within-school clustering to produce different estimates of the conditional ethnic achievement gap is not straightforward. For example, Sisben and the indicator created by prediction from linear regression explain similar proportions of the overall achievement gap between White and (all) minority students (29.3% and 31%), nonetheless, the former is a worse predictor of achievement (with adjusted R^2s of 0.05 and 0.11) and is more clustered within schools (with intras-school correlations of 0.69 and 0.37) than the latter. Similarly, the indicator created by prediction from linear regression and the indices that use standardised variables estimate a similar ethnic gap in SES (0.08 SD and 0.07 SD) and have a similar predictive power for maths achievement (with adjusted R^2s of 0.11 and 0.09) but explain different proportions of the ethnic achievement gap (31% and 24%).

The extent to which the choice of SES indicator affects the estimation of the conditional ethnic achievement gap changes by ethnic group, with the gap for Indigenous students being more influenced by this choice. Nonetheless, there were little differences between using categorical or continuous versions of the variables and all SES opera-

[18]Please note that it is not possible to compare this finding with the existing literature, since each ethnic achievement gap study reports the conditional ethnic gap after deciding on a particular operationalisation of SES; without comparing SES operationalisations. The substantive implications of the finding that SES partially explains the ethnic achievement gaps are discussed in chapter 8.

tionalisations studied here (excepting for Sisben) showed that SES is a more important explanation of the ethnic gap for Indigenous than for Afrocolombian and other minority students. This result, however, seems to be context-dependent, as Braveman et al. (2001), in turn, found that not only the magnitude but also the overall conclusions for all minority groups (African Americans, Asian/Pacific Islanders, and Latinas) were affected by the choice of continuous or categorical operationalisations of SES, in a study of ethnic gaps in health outcomes for mothers and newborns in California. This may be explained because, as noticed earlier, the correlations between indicators in Braveman et al. (2001) data are smaller than the correlations in the SABER 11 data analysed here.

Other aspect that is unique in this chapter is the consideration of multiple levels of the education system (students, schools and LAs) when studying the ethnic achievement gap and how the operationalisation of SES influences it. Consistently with the analysis in chapter 6, this chapter did not only consider the overall ethnic achievement gap, as in the existent literature, but also the within-school gap and the school and LA contextual effects of ethnicity. Section 7.4.4 showed that studying these components of the gap requires more careful consideration of the operationalisation of SES, as it is more likely to arrive at contradictory results in comparison to the analysis of the overall gaps. For example, operationalising SES using Sisben would lead to the conclusion that Afrocolombian and the group of other minority students are more socioeconomically disadvantaged than their White peers attending the same school, but all other indicators lead to the opposite conclusion.

These findings show that there is no 'ideal' SES operationalisation, as there are associated pros and cons with each choice. The prediction of linear regression method is desirable in terms of fit, but it may incorporate distortions in the analysis by combining information about maths achievement and not only education, occupation and income. Additionally, as the weights of this index are based on a regression model that fits maths test scores, it would be expected that its predictive power and ability to explain the ethnic achievement gap are lower when other subjects are analysed. In other words, the advantages of this indicator may be dependent on the outcome that is analysed, and it would be expected for it to work better for subjects that are strongly correlated with maths. In turn, the methods using standardised variables do not depend on a specific dependent variable (maths test scores) but are not as good predictors of maths achievement. Besides, it is unclear if the weights can be used in other samples (like different cohorts, for example), as this chapter did not explore the stability of the weights over time. This and other limitations of this chapter are discussed in the next section.

7.5.2 Limitations

Despite contributing to the achievement gaps literature by identifying and comparing the index creation methods that are commonly used in this research area, this chapter is limited in some ways. First, as it focuses on the methods outlined in the systematic literature review of academic journals that are indexed in Scopus (which, in theory, reflect current research practice), it does not provide an exhaustive examination of all possible index-creation methods. The main drawback of this limitation is that current practice does not necessarily reflect the best possible practice. For example, instead of transforming the original variables to undertake PCA, an option is to use its categorical-variable equivalent Multiple Correspondence Analysis, and the same might be said for the categorical-variable equivalents of FA, which also has extensions to multilevel data as explored by Yang (2003); Yang and Gustafsson (2004); Yang Hansen and Munck (2012). Although this chapter showed that very similar results are achieved when using PCA and FA, which also use information of the covariance matrix to assign weights to the indicators, whether there are practical reasons to prefer these more sophisticated methods over the ones studied in this chapter in the context of achievement gaps is still a matter of discussion.

Second, this chapter has been limited to the use of education, occupation and income as input variables for the indices, given the consensus in the literature about what variables constitute SES. However, it does not necessarily reflect how the methods are used in the literature. For example, the PISA index of socio-economic and cultural status, which uses PCA, includes home possessions instead of income. It has been argued that possessions may be a more reliable and easy-to-collect indicator of access to resources than income (APA Task Force on Socioeconomic Status, 2007; Buchmann, 2002; Cowan et al., 2012). However, there is no evidence about the differences between using income and home possessions when creating SES indices through different methods.

Another way in which the application in this chapter differs from relative common practice is in the use of the occupational prestige scale. Commonly, stay-at-home parents and pensioners are treated as missing values, while here these were assigned a prestige of zero. This was done to avoid an additional level of complexity that would be the treatment of missing data and to be able to compare the index creation methods against using the original observed variables. Further research is required to understand the consequences of using different index-creation methods with non-at-random missing data to measure SES when studying the achievement gap. Although G. N. Marks (2011) showed that changing the method for handling missing data is not associated with substantial changes in the correlation between SES and achievement, this chapter showed that this is only one of the possible mechanisms through which this methodological decision may change the conclusions of the analysis. Additionally, this chapter does not examine the consequences of other common practice for researchers: splitting continuous SES indicators into categories. The topic of discretisation, however, has been already explored and there is a consensus in the recommendation of not splitting continuous variables (e.g. Gelman & Park, 2009; Royston, Altman, & Sauerbrei, 2006).

A third limitation of this chapter is that it does not discuss the details of using FA and PCA, including rotations or the number of factors/components that would be optimally extracted from the data. This is because, regardless of the number of factors that would be extracted in an exploratory analysis, the practice is to use the first component of the data to represent SES. Besides, the focus of the chapter is on the consequences of using the indicators, and not on the dimensions that could be extracted using these kinds of analyses. This has been discussed elsewhere (e.g. Bartholomew et al., 2008; Costello & Osborne, 2005). Appendix A.5 presents additional details about these methods and the results of the analysis for the SABER 11 data.

A fourth limitation is that this chapter did not analyse the issue of measurement invariance and therefore assumed that education, occupation and income combined in the same way to represent SES for all ethnic groups. Lenkeit et al. (2015) and G. N. Marks (2011) have shown that these variables do not represent the same construct for all ethnic groups in the UK and countries around the world, respectively. In this chapter, the decision of using the same weights for the composite indicators for all ethnic groups allows examining whether the same indicator has a different effect on the results for each ethnic group without the influence of differential weight for each of them. The findings show that this is the case indeed, and future research can explore if this is explained by the lack of measurement invariance among ethnic groups.

A fifth limitation of the analysis in this chapter is that it is based on Colombian data for the 2008 cohort. It does not examine how to extend the findings to other cohorts or whether the methods differ in their stability across time, nor how this may affect the results of a study about achievement gaps. Future research should address this point, especially focusing on the prediction from linear regression method. Chapters 8 and 9, however, show that the results are fairly consistent over time when using the sum of standardised variables to operationalise SES.

Finally, although the steps for examining competing index creation techniques may be replicated in other settings, this chapter has limited its focus to research about ethnic achievement gaps. When the gap of interest is not between ethnic groups, other outcomes should be under study, most evidently, the estimated SES achievement gap.

These last two limitations arise from the fact that it is not possible to observe SES, as this is a latent variable. Given this limitation, it is not possible to assess whether the baseline for comparison (including 'the big three' as control variables) provides closer estimates of the true parameters. These parameters include the distribution of SES among ethnic minorities, its clustering within schools and LAs and its relationship with academic achievement. A way of addressing this limitation is to undertake simulation studies that allow understanding what the population parameters that would lead to the results that are observed in the application are. The challenge then is to create simulated populations that apply to a large enough number of applications.

7.5.3 Implications for Future Research

This chapter's implications arise from both its findings and limitations. The main recommendation for researchers is to be mindful that the use of composite indicators may incorporate some distortions into the analysis and to be clear about how the analysis may be affected by their inclusion. Since there is no operationalisation of SES that is preferable over others under all criteria and the context of the research is likely to play an important role in this decision, to accomplish this, the chapter provides a framework for analysing how such weights may affect the findings of the results for the case of ethnic achievement gaps. This framework includes the analysis and reporting of the distribution of indicators among ethnic groups, the predictive power, clustering patterns and the conditional gaps at each level of the school system (student, schools and LAs) and overall. Alternatively, the transparency of the findings could be increased by reporting the estimated parameters of interest (the conditional ethnic achievement gap, for example) after using competing measures of SES, when possible.

In line with the existing literature, when picking a method for measuring SES, it is also important to be clear about the objective of the research, the purpose of controlling for SES and the research context (APA Task Force on Socioeconomic Status, 2007; Braveman et al., 2001; Cowan et al., 2012). If the aim is to obtain results that are as similar as possible to those obtained when including all observed variables in the regression model, prediction from linear regression is the most appropriate option at the cost of incorporating information about the predicted variable (in this case maths test scores). The indicators created using standardised variables are also highly clustered within-schools, although not as much as the indicators created by linear prediction, but have a lower predictive power, which is reflected in the conditional ethnic achievement gaps. The recommendation would be to prefer any of these methods or the use of Sisben or an equivalent indicator used for policy-allocation purposes (such as FSM in the UK) to the sum of unstandardised variables and the Hollingshead (2011) four-factor index, since they aggregate variables with different scales and are weakly correlated with achievement. However, if including all the SES variables in the model does not generate multicollinearity problems (as is the case of this application), this would be preferred as it avoids these kinds of distortions. If multicollinearity is a problem, it is preferable to use composite indicators instead of excluding education, occupation or income, as all these variables are relevant predictors of maths achievement. This last recommendation is in line with Johnston et al. (2018)'s in the context of studies for voting behaviour.

For database administrators, the recommendation is to include as much information as possible in its original form of collection. This will allow researchers to create composite indicators according to their research needs and study how their choices may influence their conclusions. One database may be used to answer a large number of research questions and testing each composite indicator to make sure it is suitable for every research question is impossible.

These recommendations not only apply to the analysis of SES but also to other latent variables, including, for example, student attitudes and motivation which, although not as widespread as SES, have been highlighted as important predictors of academic achievement and are usually measured through a variety of methods (White, 1982; Willms, 1992).

For the next chapters of the thesis, SES will be measured as the average of standardised variables. This is a more parsimonious method that can be easily adapted to the analysis of different cohorts without adding further distortions, although at the cost of some explanatory power.

Regarding future research, it is necessary to explore a broader range of index-creation methods, including latent trait and latent class analysis, in order to understand how these distortions may be minimised. Possibly this comparison of methods could be undertaken using simulation methods instead of specific applications, but focusing on the outcomes that are relevant for this kind of applications, such as the conditional ethnic achievement gap or the degree of clustering. Future research can also examine which indicators perform better when there are missing data problems and investigate if the assumption of measurement invariance holds for the Colombian context.

7.6 Summary

This chapter has contributed to the achievement gaps in the literature by:

1. Providing an overview of the index-construction techniques that are used in the achievement gap literature to operationalise SES, which include FA, PCA, the average and sum of standardised variables, the sum of unstandardised variables, the Hollingshead (2011) four-factor index and prediction from linear regression.

2. Exploring how different indices perform, instead of exploring different variables to include in the index as previously in the literature (APA Task Force on Socioeconomic Status, 2007; Cowan et al., 2012; Ilie et al., 2017; G. N. Marks, 2011; Sirin, 2005; White, 1982). This showed that the approaches for operationalising SES can be divided in five different categories of operationalisations, which have similar distributions among ethnic groups, predictive power, level of within-school clustering and estimate similar conditional ethnic achievement gaps to other operationalisations within the same category, but not to other categories. These categories are:

    - Including individual variables (either as continuous or categorical variables)
    - Using a measure that is already used for public policy (Sisben in Colombia, which is comparable to FSM status in the UK).
    - The inclusion of all variables into the regression model and an index created by prediction from linear regression.

- Indices created using standardised variables (those created by FA, PCA and the average and sum of standardised variables).
- Indices that do not use standardised variables (the Hollingshead (2011) four-factor index and the sum of unstandardised variables).

3. Incorporating the analysis of different levels of the education system (students, schools and LAs) to consider how they are affected by the operationalisation of SES and in turn, combine to influence the overall conditional ethnic achievement gap. This showed that how these components combine is not straightforward and varies by ethnic group, with the operationalisation of SES having a stronger influence on the conditional achievement gap between White and Indigenous students.

4. Providing a framework for the analysis of the influence of the operationalisation of SES on the findings of an ethnic achievement gap study, which intends to increase the transparency of the conclusions of future research.

This chapter has also informed the decision of the operationalising SES as the average of standardised variables for the next chapters of this thesis, which use SES to explore possible explanations of the ethnic achievement gap in Colombia and whether the ethnic achievement gap varies for different subgroups of students, schools and LAs, including those with different SES. The advantages and disadvantages of using this index are discussed in sections 7.4 and 8.4.

8 To What Extent do Student, School and LA Characteristics Explain the Overall Ethnic Achievement Gaps and their Components at each Level of the Education System? How do Single-Level and Multilevel Modelling Results Compare when Answering this Question?

8.1 Introduction

Chapter 6 presented the estimation results for the unconditional achievement gaps between White and minority 11th grade (age 16) Colombian students. After showing that the ethnic achievement gaps are a combination of ethnic differences within and between schools and local authorities (LAs) with a different ethnic composition, chapter 6 pointed out the problem by showing that these last two accounted for most of the gaps between White and minority students. This result contrasted with the findings in the US and the UK, where differences between schools are less important, and led to the broad policy recommendation of prioritising interventions at the LA and school-level. As the literature review (section 2.7) and methods (sections 5.2 and 5.3) chapters explained, since the overall ethnic achievement gap is an average difference, it results from the combination of many other variables that are linked to the students' ethnicity and academic achievement. This chapter examines what these variables are.

One of these possible variables is socio-economic status (SES), as explored in chapter 7, which showed that Afrocolombian and Indigenous students have a lower average SES than White students and that students with higher SES score higher on the SABER 11 maths test. As a result, SES[1] explains around 18.2% of the overall maths achievement

[1] Measured by the average of standardised parental occupation and education and family income.

gap between White and Afrocolombian students, 47.9% for Indigenous students and 1.1% for other minority students. Therefore, chapter 7 showed that a third variable (SES) changes the relationship between ethnicity and maths achievement (the ethnic achievement gap). However, this was the only variable that was considered.

This chapter explores whether student (gender, SES, household size and age), school (type: state or private, zone: urban or rural, day type: full day, morning or afternoon, focus: academic, technical or teaching, size and adherence to an ethnoeducation-based program) and LA (availability and management of resources and the incidence of conflict and violence) characteristics explain the overall and within-school ethnic achievement gaps, and the school and LA contextual effects of ethnicity between 2008 and 2013.

In the international literature, it is common for authors to consider variables such as age, gender, SES and school climate and composition when studying ethnic achievement gaps to try to estimate the effect of ethnicity net of the effect of these variables (e.g., Bali & Alvarez, 2004; Condron, 2009; Dustmann et al., 2010; Grogan-Kaylor & Woolley, 2010; Stiefel et al., 2007; Wilson et al., 2011)[2]. However, it is not common to examine the extent to which these characteristics explain the gap. As explained in the methods chapter (section 5.3) this kind of analysis is performed using either the Oaxaca (1973) decomposition or mediation analysis.

In Colombia, Sánchez-Jabba (2011) used the Oaxaca (1973) decomposition to show that student and school characteristics explained around 60.6% of the overall maths achievement gap between White and minority students (pooled as a single group) in 2010, as summarised in section 3.4.2. However, given the limitations of his data and methodological approach, he neither studied how LA characteristics influenced the achievement gap, whether there were different patterns across ethnic minority groups (Afrocolombian, Indigenous or other minorities), how these variables interrelated at different levels of the education system (students, schools and LAs), nor checked if the results were persistent over time. The Oaxaca (1973) decomposition has also been used around the world to differentiate what proportion of a gap is explained by the observed variables and what is not (e.g., Arteaga & Glewwe, 2019; Blanco, 2017; Cook & Evans, 2000; Santibañez, 2016).

Mediation analysis has been used to show that ethnic differences in anxiety (Osborne, 2001), test preparation (Dollinger & Clark, 2012; Ellis & Ryan, 2003) and physical

[2]Some of these authors also explored interaction effects, which are a matter of discussion in chapter 9.

activity (Masel, Raji, & Peek, 2010) partially explain the overall ethnic achievement gap, while multilevel mediation analysis has been used to show that teaching quality (Rjosk et al., 2014), school disciplinary climate and the principals' perception of students' behaviour and morale (Liu et al., 2015) mediate the relationship between the schools' ethnic composition and academic achievement. Nonetheless, these methods have not been used to simultaneously study the components of the gap (the within-school gap and the school and LA contextual effects). This is probably the case because the fact that single-level coefficients also incorporate information from different levels of a hierarchical structure (such as students nested within schools, as discussed in chapter 6) has not received substantial attention in the applied literature. Besides, the use of multilevel modelling is often applied uncritically when analysing hierarchical data, without considering alternative methods.

Nonetheless, as discussed in section 5.2 in the methods chapter, multilevel modelling is not the only approach to examine how different levels of the education system contribute to explain the ethnic achievement gaps. The main alternative is to use single-level (Ordinary Least Squares (OLS)) regression, which was the approach taken in chapter 6 to decompose the ethnic achievement gap into its components at each level of the school system. Multilevel regression has the advantage of directly modelling complex data structures to estimate standard errors that account for the data dependencies in such structures, unlike single-level regression. However, multilevel modelling estimates of overall effects are shrunk towards the within-group or between-groups effects, depending on how similar the members of the group are to each other and to members of other groups, which may distort the analysis if the parameter of interest is the overall mean difference (Bell & Jones, 2015; Rabe-Hesketh & Skrondal, 2012). For this reason, this chapter also discusses the advantages and disadvantages of using single-level and multilevel model analysis when studying how different student, school and LA characteristics explain the ethnic achievement gaps and their components.

As discussed in section 3.4, quantitative educational research in Colombia has extensively searched for variables that can predict academic achievement, in a body of literature that is known as 'associated factors' (*factores asociados*). Thanks to works in this area, there is a consensus about the importance of gender, SES, school type (state or private) and the type of school day[3] as predictors of academic achievement (Barrientos

[3] Please refer to section 3.3.3 and chapter 4 for further details about these variables.

Marín & Rios, 2007; Bentaouet Kattan et al., 2008; Correa, 2004; Gaviria & Barrientos Marín, 2001a, 2001b, 2001c; Martínez Barrera, 2012; Piñeros Jiménez & Rodríguez, 1998). These variables potentially explain some of the achievement gaps if they are differentially distributed among ethnic groups; for example, if gender differences in enrolment are narrower for any ethnic group, or if minority students disproportionally attend state schools and are therefore analysed in this chapter. Nevertheless, since the study of ethnicity is not prevalent in quantitative educational research in Colombia, one addition to the list of characteristics is whether school curriculum is based on ethnoeducation (a special education program for ethnic minorities, described in section 3.3.1.2). While the existence of a special type of education for ethnic minorities is national policy since the mid-1980s (Calvo Población & García Bravo, 2013; Castillo Guzmán, 2008), there is no evidence about how it relates to academic achievement, as measured by the SABER 11 exams, or how it relates to the ethnic achievement gaps.

To reach its objectives, the remaining of this chapter is organised as follows:

1. Section 8.2 reviews the studies that use observational data to examine possible explanations for the overall ethnic achievement gap and its components (the within-school gap and the school and LA contextual effects of ethnicity).

2. Section 8.3 presents the estimated models and discusses their interpretation.

3. Section 8.4 presents the variables that are included in the analysis.

4. Section 8.5.1 evaluates the advantages and disadvantages of using single-level and multilevel models to analyse to what extent student, school and LA characteristics explain the ethnic achievement gaps.

5. Sections 8.5.2 to 8.5.5 examine the influence of student, school and LA characteristics on the overall ethnic achievement gaps and their components (the within-school ethnic achievement gap and the school and LA contextual effects of different ethnic groups) in Colombia between 2008 and 2013.

6. Section 8.6 discusses the findings, their limitations and the implications for policy and future research.

7. Section 8.7 summarises the contributions of the chapter.

8.2 Literature Review

As discussed in the general literature review (chapter 2), there is an extensive body of literature that aims to explain the existence of ethnic achievement gaps around the world (e.g. Fryer & Levitt, 2004; McEwan, 2004; Strand, 2011). Within this literature, student, school, LA and country characteristics are proposed as possible explanations of the achievement gaps, with a focus on the overall ethnic achievement gap but not on how these variables explain each of its components (the within-school gap and school and LA contextual effects). What is more, authors do not often recognise the within-school gap and the contextual effects as part of the overall gap but interpret them separate phenomena (such as peer effects). Section 8.2.1 discusses the studies that have aimed to explain the overall ethnic gap, while section 8.2.2 focuses on those that have tried to explain the components of the gap.

Section 8.2.3 summarises the methodological discussion regarding the choice between single-level and multilevel modelling when examining how different variables explain the components of the gaps, simultaneously.

8.2.1 Explaining the Overall Gap

As explained in the methods chapter (section 5.3), there are two equivalent methods to study the role of student, school and LA characteristics in explaining the ethnic achievement gap. The first one is using the Oaxaca (1973) decomposition, and the second one is examining how the (unconditional) gap changes when including control variables (as in mediation analysis).

The Oaxaca (1973) decomposition is standard in labour economics and can be used to split the ethnic achievement gap into two parts: one that is explained by observed characteristics and another one that is not. In Colombia, Sánchez-Jabba (2011) used the Oaxaca (1973) decomposition to show that working status, gender, household size, zone (rural or urban), family income, mother's education, school-day type and school fees explained 60.6% of the 0.35 standard deviations (SD) overall maths achievement gap between 11th grade (age 16/17) White and minority students in 2010.

The same approach has been applied in different countries, also with limited success in explaining the gaps. For example, in Mexico, Blanco (2017) estimated that differences in SES, aspirations and school type explained 72.4%, of the 0.29 SD maths achievement gap between non-Indigenous and Indigenous students at the end of 6th grade (age 11) in 2013. He concluded that further research was required to identify the factors that they could not observe and could potentially explain the remaining part of the gap.

Page et al. (2008) combined the Oaxaca (1973) decomposition technique with Reardon (2008)'s decomposition into a within-school, between-school and ambiguous components of the gap, discussed in detail in chapter 6. Page et al. (2008) used this combined method to examine how parental education, gender and exposure to other ethnic groups could help explain the overall achievement gap in maths for grades 4, 8 and 11 (ages 9/10, 13/14 and 16/17) students in the US between 1978 and 2004. Page et al. (2008) found that parental education explained between 17.2% and 24.2% of the overall maths achievement gap, while gender and exposure to other ethnic groups did not explain more than 2.4% of the gap. Therefore, although Page et al. (2008) are part of the decomposition debate in section 6.2.1, they did not consider how the components of the gap were explained by observed variables, as it is also the case for the other studies in the debate (Cook & Evans, 2000; Fryer & Levitt, 2004, 2006; Hanushek & Rivkin, 2006; Reardon, 2008).

The second method, comparing the unconditional and conditional gaps, is used in a broader range of disciplines. Blanco (2017) complemented his analysis for Mexico using the Oaxaca (1973) with this second approach. As expected, the results he obtained using both approaches were similar, since these methods are analogous (Hou, 2014), as further discussed in the methods chapter (section 5.3).

Quinn (2015b) focused on the role of SES in explaining ethnic achievement gaps in the US. He estimated that gender, previous school experience, SES, number of children books at home, mother's age at the time of first birth and participation in a supplemental nutrition program explained around 74.1% and 67.1% of the 0.54 SD and 0.7 SD overall maths achievement gaps between White and Black and Hispanic kindergarteners in 2011, respectively. While Quinn (2015b) decomposed the Black-White achievement gap following Reardon (2008)'s methodology, he did not study how this set of controls or school characteristics affected the within-school or between-school parts of the achievement gap.

In England, Strand (2011) found that the student's sex and SES explained 78.4%, 65% and 25.8% of the 0.46 SD, 0.45 SD and 0.49 SD achievement gaps between White and Pakistani, Black African and Black Caribbean students, respectively, at the end of Year

9 (age 13/14) in 2004. Observing that SES only explained a small proportion of the gap for Black Caribbean students, in comparison to other minorities, Strand (2011) examined if differences in students' attitudes and aspirations, and parental practices explained an additional proportion of the gap. Strand (2011) showed that this was not the case since minority students tended to have more favourable aspirations and future plans, self-concept, attitude to school and fewer behavioural problems; and their parents also showed more favourable motivations, practices and school involvement, in comparison to their White peers. Despite all these favouring conditions, Pakistani, Black African and Black Caribbean students scored on average 0.24 SD, 0.41 SD and 0.38 SD lower than their White peers with similar characteristics, respectively. This implied that, jointly, all observed variables could only explain 47.1%, 7.3% and 23.3% of the gaps for these ethnic groups, correspondingly. Strand (2011) concluded that in the case of Black Caribbean students, discriminatory school policies and practices require further examination.

Unlike these studies, the focus of this chapter is not only on how these characteristics explain the overall achievement gap but on how the association between these characteristics at each level of the education system (students, schools and LAs) contribute to explain the overall ethnic achievement gap. The next section explores some studies that have examined these components.

8.2.2 Explaining the Components of the Gap

Section 5.2 in the methods chapter showed that there are different approaches for estimating the components of the gap. Including school fixed effects or estimating a model that includes the schools' ethnic composition results in the estimation of the within-school gap.

Cook and Evans (2000) were the first ones to include school fixed effects in the Oaxaca (1973) decomposition to explore possible explanations of for ethnic achievement gaps. Nonetheless, they did not decompose the achievement gap but its change between 1970 and 1988, finding that only 1.6% of the 0.18 SD decrease in the Black-White achievement gap for 11th grade (age 16/17) students in the US was explained by improvements in school quality for Black students.

This approach has been replicated in several different countries. In England in 2007, Dustmann et al. (2010) argued that 55.8% and 29.9% of the overall gap for Black Caribbean (0.16 SD) and Chinese (0.52 SD, favouring Chinese) year 11 (age 15/16) students were attributed to differences within schools rather than between schools, highlighting the importance of future research at the school level. In Latin America, McEwan and Trowbridge (2007) estimated that 0.26 SD (66%) of the 0.39 SD achievement gap between Indigenous and non-Indigenous 6th grade (age 11/12) in 2001 in Guatemala was explained by between-school differences. Similarly, McEwan (2004) estimated that 0.22 SD (55.4%) of the 0.4 SD achievement gap between Indigenous and non-Indigenous 8th grade (age 12/13) in Chile in 1997 and 0.25 SD (72.5%) of the 0.34 SD achievement gap for 6th grade (age 11/12) in Bolivia in 1997 was explained by differences between schools. As discussed in the first research chapter (section 6.6), Arteaga and Glewwe (2019) presented a version of the Oaxaca (1973) decomposition that considered community effects when decomposing the achievement gap in Peru in 2001. They added community fixed effects to show that at age 5 (before school), 89% of the ethnic achievement gap between Indigenous and non-Indigenous students was explained by differences in the effect of student characteristics, but at age 8 (after three years of school) 58% of the gap is explained by differences in community effects. Since all these studies included school fixed effects, they were unable to specify what school characteristics explained the gap.

In their study for rural China, Yang et al. (2015) found that student characteristics[4] only explained 20.7% of the 0.29 SD overall achievement gap between fourth-grade (age 9/10) minority students and the Han majority group in 2012. In turn, the same characteristics explained 40% of the 0.1 SD within-school achievement gap. Using these findings, they recommended improving pedagogical practice in such a way that benefits minority students within schools and further examining the role that language differences play.

In Minnesota, US, Myers et al. (2004) argued that student and school characteristics explained around 38.2% of the 24.54 points[5] maths achievement gap between 8th grade (age 12/13) Black and White students in 1999. Nonetheless, the schools' ethnic composition was among the characteristics that they included in the model. Therefore, theirs was a contextual effects model, and the 38.2% reduction of the gap that they reported is the result of comparing the overall achievement gap with the within-school conditional

[4]These characteristics included gender, being a boarding student, age, household size, commuting time, parental education and occupation and a set of family assets.

[5]The standard deviation for the test scores used by Myers et al. (2004) was not reported in this paper or other available documents for this period.

gap.

Fryer and Levitt (2004, 2006) tested if the importance of student characteristics increased over time by comparing the unconditional and conditional within-school gaps, after controlling for gender, age at kindergarten, birthweight, SES, number of children-books at home, mother's age at the time of first birth and participation in a supplemental nutrition program. Fryer and Levitt (2004, 2006) found these student characteristics jointly explained around 85% of the 0.64 SD within-school Black-White maths kindergarten and 57% of the 0.88 SD within-school gap in third grade. For both hypotheses, they used a model with school fixed-effects and therefore, their focus was only on explaining the within-school gap. They did not test how these characteristics explained the between-school achievement gap or the role of any specific school (or school district) characteristics in explaining the between-school achievement gap.

The models estimated in this chapter are more commonly found in the peer-effects literature, where contextual-effect models are estimated to examine the effect of the school ethnic composition. However, the interpretation of the results within the peer-effects framework is slightly different from the one in this chapter. This chapter builds on the achievement gap decomposition in chapter 6 to interpret the parameters. The most significant difference is that this chapter recognises that, when controlling for ethnic group composition, the parameter associated with the students' ethnicity is the within-school achievement gap. In turn, the peer-effects literature usually interprets this parameter as the achievement gap, conditional on group composition. Consequently, when the within-school gap is no longer statistically significant after including group composition in the model, the authors conclude that the achievement gap has been entirely explained. In contrast, the interpretation in this chapter implies that the achievement gap is not fully explained unless both, the within-school achievement gap and the contextual (or peer) effects are also explained.

For example, Noe et al. (2005) estimated a contextual effect model to describe the achievement gap between Indigenous and non-Indigenous fourth grade (age 8/9) students in Chile in 1999. The results of the single-level regression analysis showed that household income, parental characteristics and school type explain around 77% of the overall achievement gap. After including school ethnic composition in the model, the coefficient for Indigenous students is no longer significant. The authors interpreted this finding as their ability to explain the ethnic achievement gap in Chile fully. However, the contextual effect is still significant, which indicates that the authors were able to explain the within-school achievement gap, but the contextual effect of ethnicity (and therefore probably the between-school achievement gap, as discussed in detail in chapter 6) is still significant.

More recently, Verhaeghe et al. (2018) investigated whether Dutch language skills and group composition effects explained the achievement gap for students at risk in Flanders, Belgium. They defined students at risk as those with a low SES and from an ethnic minority background and focused on the interaction between SES and ethnicity. Using multilevel modelling, Verhaeghe et al. (2018) showed that language skills explained between 26.1% and 37.9% of the 0.49 SD to 1.06 SD gap between students 'at risk' and average/high SES Dutch-speaking students at the start of first grade (age 5/6), and between 26.7% and 100% of the gap at the end of sixth grade (age 11/12). In turn, group composition explained between 6.8% and 8.7% of the gap in first grade and between 3.8% and 11% of the gap in sixth grade. However, their discussion did not consider that, when including the group-composition variables, the gaps no longer represent the overall gap but the within-school gap.

Other studies in the peer-effects literature have tried to develop identification strategies to estimate the 'true' magnitude of peer effects, exploiting different sources of school or classroom composition variation (e.g. De Melo, 2014; Fruehwirth, 2013; Hoxby, 2000), but not estimating to what extent the effect of the school proportion of minority students is explained by its correlation with other characteristics. As stated by van Ewijk and Sleegers (2010b, p.240) in their meta-analysis: "virtually no study quantitatively examines the link between proposed causes and the taking place of a compositional effect".

8.2.3 Methodological Considerations

Unlike most of the previous studies, which use single-level models, Munk, McMillian, and Lewis (2014) also explored how a multilevel contextual effects model shed more light on the analysis of ethnic and SES achievement gaps. However, their focus was on how multilevel models allowed understanding how the within- and between-school variation in maths test scores, and not on how these methods compared when estimating the within-school achievement gap and contextual effects. Besides, they did not contrast the results of estimating the overall gaps and contextual-effects models when using single-level and multilevel analysis.

Similarly, in the same study previously discussed, Myers et al. (2004) analysed how their results changed when using single-level OLS regression or a hierarchical linear model. Although this was not the centre of their discussion, they concluded that both approaches were equivalent. Nonetheless, their comparison was based on a contextual effects model, which examined student-level and school-level relationships separately, as discussed in the methods chapter (section 5.2). In this case, these similarities between models are better understood, but this is not the case when the overall achievement gap is examined.

Chatterji (2006) also compared the estimated overall achievement gaps obtained from single-level and multilevel analysis, using the Early Childhood Longitudinal Study for students at the end of first grade in the US in 1999-2001. After showing that the estimated gaps were reasonably similar, decided to continue the analysis using multilevel modelling.

8.2.4 Literature Review Summary

The studies in this section showed that there is evidence that the overall and within-school gaps can be explained by a set of student and school characteristics, mainly related to their SES. There is also evidence that an important part of the overall gap results from the existence of contextual effects. However, there is less empirical evidence about the student and school characteristics that explain these contextual effects. Additionally, authors usually pick a method of analysis (single-level or multilevel modelling) and rarely question how the method choice influences their results. These are two gaps in the literature that this chapter aims to address.

8.3 Method

This chapter examines if the variables that have been found to predict academic achievement in Colombia, discussed in section 3.4.1, also explain the ethnic achievement gaps and whether they do so by explaining the within-school gaps or by explaining the school and LA contextual effects of ethnicity. The analysis here is exploratory and does not provide any evidence of the causal mechanisms behind the existence of the gap. In turn, it shows that the observed achievement gap results from the interrelations between achievement, ethnicity and other student, school and LA characteristics at different levels of the school system.

8.3.1 Explaining the Achievement Gaps and their Components

The starting point is the contextual effect model that allows estimating the within-school achievement gaps and the school and LA contextual effects of ethnicity, as shown in the methods chapter (section 5.2.1). This model is

$$y_{ijk} = \alpha + \beta_A^W A_{ijk} + \beta_I^W I_{ijk} + \beta_O^W O_{ijk} + \beta^{CSA}\overline{A}_{.jk} + \beta^{CSI}\overline{I}_{.jk} + \beta^{CSO}\overline{O}_{.jk} + \beta^{CLA}\overline{A}_{..k} + \beta^{CLI}\overline{I}_{..k} + \beta^{CLO}\overline{O}_{..k} + e_{y_{ijk}} \quad (8.1)$$

where y_{ijk} is the maths test score for student i, in school j and LA k, A_{ijk}, I_{ijk} and O_{ijk} are dummy variables taking the value of one for Afrocolombian, Indigenous and other minority students, respectively, and nil otherwise; $\overline{A}_{.jk}$, $\overline{I}_{.jk}$ and $\overline{O}_{.jk}$ are the school proportions of Afrocolombian, Indigenous and other minority students and $\overline{A}_{..k}$, $\overline{I}_{..k}$ and $\overline{O}_{..k}$ are the LA proportions of students of each ethnic minority group. Therefore, β_A^W, β_I^W and β_O^W represent the within-school achievement gap between White students and students of each minority group, β^{CSA}, β^{CSI} and β^{CSO} are the within-LA school contextual effects (in the remaining of the chapter, simply called school contextual effects) and β^{CLA}, β^{CLI} and β^{CLO} are the LA contextual effects of each ethnic group.

As explained in the methods chapter (section 5.3), evaluating the effect of ethnicity or its contextual effects on achievement, via a student, school or LA characteristic $\mathbf{z}$, only requires adding this variable to the regression model (8.1), obtaining

$$y_{ijk} = \alpha' + {\beta'}_A^W A_{ijk} + {\beta'}_I^W I_{ijk} + {\beta'}_O^W O_{ijk} + \beta'^{CSA}\overline{A}_{.jk} + \beta'^{CSI}\overline{I}_{.jk} + \beta'^{CSO}\overline{O}_{.jk} + \beta'^{CLA}\overline{A}_{..k} + \beta'^{CLI}\overline{I}_{..k} + \beta'^{CLO}\overline{O}_{..k} + \delta\mathbf{z} + e_{y_{ijk}} \quad (8.2)$$

After including $\mathbf{z}$, the estimated parameters for the within-school gaps and the contextual effects change -represented by the apostrophe(′)-, as they now are the within-school gaps, school and LA contextual effects *conditional* on $\mathbf{z}$. This means that the gaps are estimated after comparing students with the same characteristic $\mathbf{z}$. The change in the achievement gap (or contextual effect), calculated as the difference between the original (unconditional) gap and the conditional gap is equivalent to the indirect effect of ethnicity on achievement through the characteristic $\mathbf{z}$.

In the results, the change in the gaps is shown such that a positive sign indicates a wider gap and a negative sign indicates a narrower gap. A wider gap indicates that when comparing White and minority students with the same characteristic **z**, minority students score even lower than when randomly selecting White and minority students. A narrower gap is usually understood as evidence of an effect of ethnicity on achievement through **z** or as that **z** explains or mediates a proportion of the achievement gap (MacKinnon et al., 2007; Tofighi & Thoemmes, 2014; Z. Zhang, Zyphur, & Preacher, 2009). In both cases, the change in the gap arises as a combination of two factors: First, the characteristic **z** is correlated with maths achievement y_{ijk} and second, the characteristic **z** has a different distribution for White and minority students. For example, if there is an indirect effect of ethnicity through SES on maths achievement or, analogously, if SES explains the gap, it means that first, students with lower SES tend to obtain lower maths scores and second, minority students tend to have a lower SES.

The existence of an indirect effect (or that the gap changes after including a control variable) does not provide evidence of any direction of causality. In particular, if the focus were on SES instead of ethnicity, the change in the SES coefficient after including ethnicity in the model would be interpreted as evidence that ethnicity explains the effect of SES. This finding implies that minority students tend to obtain lower test scores and students with a low SES usually belong to an ethnic minority group. Therefore, none of the statements when presenting the results should be interpreted as providing evidence that the characteristic **z** *causes* the ethnic achievement gap. In turn, the change in the estimated gap provides additional information on the patterns that arise in the data and how they *correlate* with the gap.

Z. Zhang et al. (2009) and Tofighi and Thoemmes (2014) argued that, in the context of multilevel mediation analysis, when the **z** is a student characteristic z_{ijk}, it may affect the estimates of the within-school gaps and the school and LA contextual effects of minority students. This happens because including z_{ijk} in the model captures information not only about the student but also about how z_{ijk} is distributed among schools and LAs. In contrast, when **z** is a school characteristic z_{jk}, it will not affect the estimates of the within-school gaps, but it may change the estimates of the school and LA contextual effects if there is an indirect effect. This is because adding z_{jk} to the model does not incorporate any new information about individual students, but only about the schools and how they are distributed among LAs. For a similar reason, if **z** is a LA characteristic, including it in the model will neither change the within-school gaps nor the school contextual effects, as it only incorporates new information about the LAs. In turn, since the overall achievement gap is a combination of the within-school gap and school and LA contextual effects[6], its change reflects how the student, school and LA characteristics explain the within-school gap and the school and LA contextual effects of ethnicity.

8.3.2 Single-Level and Multilevel Modelling

The distinction between within- and between-cluster (e.g. school) effects and their relationship with mediators is relatively well-known in methodological and applied research in multilevel mediation analysis (Knifsend et al., 2018; Ndum, Allen, Way, & Casillas, 2018; Pituch & Stapleton, 2012; Wolfgramm et al., 2014, e.g.). Nonetheless, since the fact that single-level regression coefficients are also the result of a combination of within-and between-cluster effects is less well-known, the process outlined above is usually carried out by estimating multilevel models.

When the models above are single-level regression models, the assumption is that $e_{y_{ijk}} \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right)$. In turn, when the models are understood as multilevel (random-intercept) models, it is assumed that $e_{y_{ijk}} = v_k + u_{jk} + r_{ijk}$, where $v_k \overset{i.i.d.}{\sim} N\left(0, \sigma_v^2\right)$, $u_{jk} \overset{i.i.d.}{\sim} N\left(0, \sigma_u^2\right)$ and $r_{ijk} \overset{i.i.d.}{\sim} N\left(0, \sigma_r^2\right)$.

As further discussed in section 5.2, multilevel modelling produces precision-weighted coefficients that differ from those estimated from single-level models depending on how different the school and LA sizes are (how unbalanced the data are) and on the dispersion of maths test scores within and between clusters.

Multilevel modelling is usually thought of as the gold standard for handling data with a nested structure, and little consideration is given to other options. One alternative is using single-level regression modelling with cluster-robust estimates of the standard errors of the models' slopes and intercept. As discussed in section 5.2, there are trade-offs when choosing between single-level and multilevel models. Multilevel modelling provides better estimates of the standard errors of intercepts and slopes, which facilitates inference about the achievement gaps. However, the multilevel modelling estimate of the overall gap is a precision-weighted average of the within-school achievement gap and the school and LA contextual effects of ethnicity that can be hard to interpret (Bell & Jones, 2015; Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002).

[6] Please refer to chapter 6 for a detailed discussion.

These estimates will be closer or further from the OLS single-level-model estimate of the achievement gap depending on how clustered the data are at each level and the uncertainty associated with each estimated parameter. As shown in the data chapter (section 4.4) and section 7.4, in Colombia, students within schools tend to be very similar to each other and different from students in other schools. Hence, as further explained in section 5.2.2, the multilevel estimate of the overall gap is expected to be closer to the within-school gap than to the between-school gap. However, given that this is context-dependent, how different single-level and multilevel models' estimates are is an empirical question.

When estimating the within-school achievement gap and school and LA contextual effects of each ethnic group, multilevel models provide unbiased estimates of these parameters. Besides, these models penalise (shrink) those observations with high uncertainty and, therefore, the estimated changes in the gaps will be less likely to be distorted by unusual observations. In turn, single-level estimates of the change in the gap are more likely to be distorted by schools or LA with unusually high values on any of the characteristics. Additionally, single-level regression requires the estimation of consistent standard errors when analysing the school and LA contextual effects, which is difficult when the hierarchical structure includes more than two levels or when the nature of nesting is more complex (Cameron, Gelbach, & Miller, 2011).

Since researchers usually pick either single-level or multilevel regression without further comparison, little is known about how the choice between single-level and multilevel regression analysis impacts the results of this kind of analysis. This chapter contributes to understanding this difference by examining how the results that both kinds of models produce differ in this particular application.

8.4 Data

This chapter analyses the SABER 11 dataset (described in chapter 4) for the cohorts between 2008 and 2013. The variables under analysis match with the variables that have been used in the literature, as discussed in section 3.4, with two notable exceptions. First, the inclusion of variables about schools' infrastructure and educational resources is not possible as these are not available for any of the years in the sample. Second, the literature does not usually consider how ethnoeducation programs correlate with student achievement, but this chapter explores if this variable contributes to explaining the ethnic achievement gap.

These variables are:

- Student characteristics:
  - Gender
  - Household size
  - Age
  - SES

    The measure of SES is the average of standardised parental education, occupation and income that is further described in chapter 7. The reason to choose this index is that it offers the advantage of using a single measure of SES that can be easily constructed over time without incorporating the additional distortions that would arise from calculating different weights for each year or using the same set of weights for all years.

    If alternative methods for constructing a SES index were used, the observed changes in the gaps could be the result of either different weights for each year (if different sets of weights are estimated for each year) or changes in the fit of the weights for each of the SES variables (if the same sets of weights are used for each year). That is, changes in the proportion of the gap that is explained by SES over time would not only reflect changes in the ability of SES to explain the gap, but also changes in the measurement of SES. As discussed in chapter 7, the properties of the index created as the average of standardised SES variables are similar to the properties of indices created by factor analysis (FA) or principal component analysis (PCA). When compared to including all individual SES variables in a regression model, the average of standardised SES variables tends to explain a smaller proportion of the achievement gap and the model fit tends to be poorer. However, these disadvantages are compensated by the ability to summarise all SES variables into a single variable, which improves the parsimony of the model.

- School characteristics:
  - Type (state or private)

- Zone (urban, rural or both[7])
- Day type (full day, morning or afternoon)
- Focus (academic, technical, teaching and academic and technical)
- Whether or not the school follows a pedagogical approach for ethnic education (ethnoeducation)[8]
- Number of students from each school who took the SABER 11 exam each year, used as a proxy of school size.

- LA characteristics:
  - Fiscal conditions[9]
    * Transfers for running costs
    * Transfers for quality
    * Fiscal performance.
  - Violence[10]
    * Pressure
    * Intensity

      From two sources: the actors of the inner conflict and organised crime. When the perpetrators are inner-conflict actors, the variable name includes 'conflict' ('Conflict Int' and 'Conflict Pres' for intensity and pressure, respectively), while when the perpetrator is organised crime, the variable name includes 'crime' ('Crime Int' and 'Crime Pres').

  Some studies have considered additional variables, such as aggregated poverty or living-condition indicators (García Villegas et al., 2013; Mina Calvo, 2004). In this chapter, the inclusion of the student SES in the model (which has not been school- or LA-mean centred) implicitly controls for the LAs' SES composition, and therefore provides similar information about the LAs. In other

[7]Schools with more than one venue may have venues in both, urban and rural settings. Students may take classes in only one or many venues of the school.

[8]For a description of this teaching approach, please refer to section 3.3.1.2.

[9]Please refer to sections 3.3.2.1 and 4.5 for further information about resource allocation in Colombia and how it is distributed among LAs with different ethnic composition, respectively.

[10]Section 3.2 presents and overview of the sources of violence in Colombia and section 4.5 describes how it is distributed across the country.

words, including student and school-level characteristics into the model implies that the models control for the LA composition for all these characteristics.

All the continuous control variables have been standardised for each year to have a mean of zero and variance one, which facilitates the comparison when analysing the changes in the achievement gap.

8.5 Results

Including control variables into the model means that the estimated achievement gaps now represent the average maths score difference between White and minority students with the same characteristics, attending schools and LAs with the same characteristics (including their composition). The same happens when examining the components of the gaps (the within-school gaps and the school, and LA contextual effects of ethnicity).

Including only one predictor at a time allows examining which of the predictors explains the largest proportion of the gap or contextual effect at each level. Including all the predictors at the same time allows estimating the proportion of explained gap net from all interrelations between these variables. As explained in the methods chapter (section 5.3), in a mediation analysis framework, this is equivalent to estimating the indirect effect of an individual mediator or including multiple mediators in the model.

The next section compares the estimates of the proportion of the explained gaps and their components obtained from single-level and multilevel models. After discussing the advantages and disadvantages of each method, and how these influence the interpretation of the results, the estimated proportion of the gaps that are explained by these characteristics are presented.

8.5.1 Estimating the Conditional Gap Using Single-level and Multilevel Models

As discussed in sections 5.2.2 and 8.3, multilevel modelling provides precision-weighted estimates of regression parameters. This feature can be an advantage, but also a disadvantage, depending on the parameter of interest. This section shows the multilevel and single-level model estimates of the proportion of the gap that is explained by different student, school and LA characteristics. Precision-weighting may be a disadvantage, as the multilevel estimates of the overall achievement gaps are shrunk towards the within-school achievement gap. Simultaneously, the precision weighting is an advantage, as it prevents the estimates of the within-school gaps and the school and LA contextual effects from being distorted by unusual observations. The discussion in this section is limited to the methods comparison for the set of all student, school or LA characteristics presented in section 8.4. Later sections focus on individual characteristics.

8.5.1.1 Estimation of the Overall Gaps

Figure 8.1 shows the unconditional ethnic achievement gaps for each ethnic group, as estimated by single-level (SLM) and multilevel models (MLM). As this figure shows, the overall achievement gaps estimated through multilevel and single level modelling are very different, since the multilevel modelling achievement gaps are closer to the within-school gaps. However, how different the estimated gaps are varies by ethnic group. For Afrocolombian students, the difference is larger, with the single-level models estimating an overall gap between 0.42 SD in 2013 and 0.59 SD in 2011, while multilevel models estimate an overall unconditional gap of between 0.08 SD in 2012 and 0.14 SD in 2008 for this ethnic group. In contrast, for the group of other minorities these estimated parameters lie between 0.1 SD in 2008 and 0.2 SD in 2012 for single-level model estimation of the overall gap, and between 0.03 SD in 2009 and 0.08 SD in 2012 for the multilevel model estimation of the ethnic achievement gap.

These differences are also reflected in the estimations for the proportions of the gaps that are explained by different student, school and LA characteristics. Figure 8.2 shows the percentage change in the overall achievement gap, after including the set of controls indicated by different shades (all student-level variables, all school-level variables, all LA-level variables or all control variables). The figure contrasts the methods used to estimate the gaps (single-level models or multilevel models). A positive change in the gap means that the achievement gap increased after including controls, while a negative change means that the estimated achievement gap decreased after including controls[11].

As this figure shows, single-level models report a reduction in the overall achieve-

[11] Under standard mediation analysis, including mediators should result in a reduction of the achievement gap and therefore the convention would be to report the absolute value of this change (as the indirect effect, as measured here, would always have a negative sign). In this thesis, an increase in the achievement gap is also of interest, and therefore the decision was to keep the signs to indicate an increase or a decrease in the estimated achievement gap.

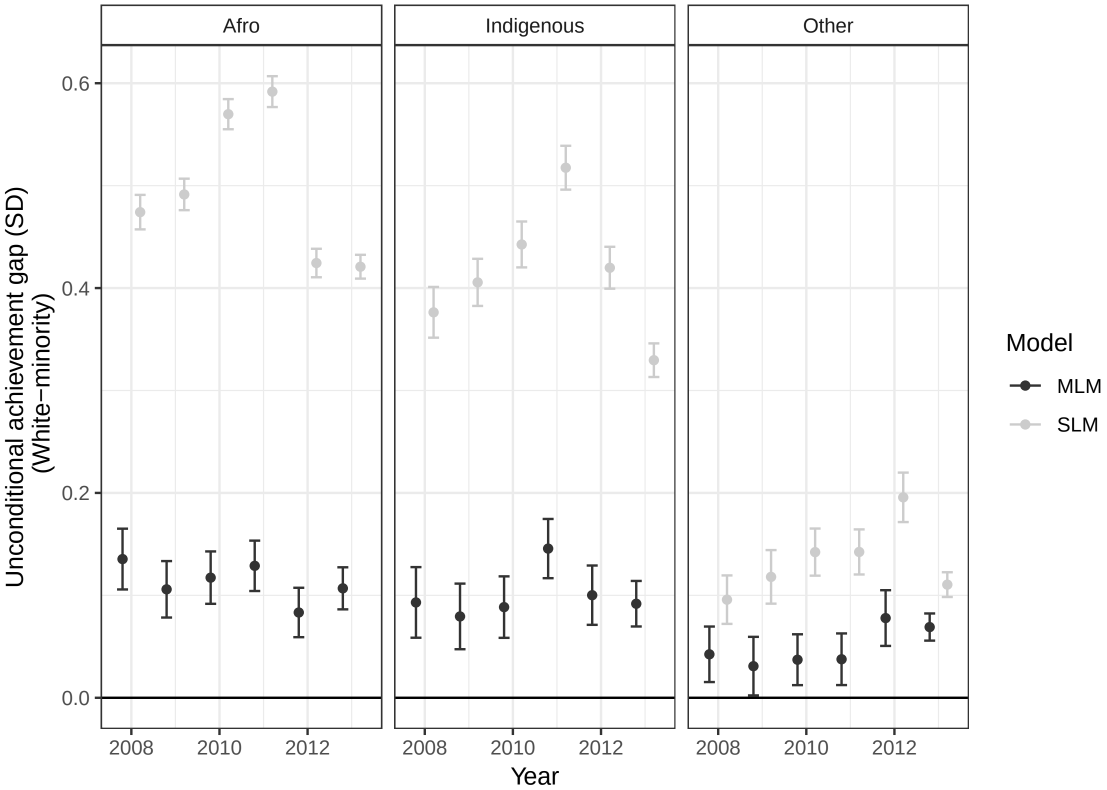

Figure 8.1: Estimated unconditional overall ethnic achievement gaps by estimation method

ment gap after considering student, school and LA characteristics (as different sets and simultaneously) for all ethnic groups. In fact, according to the single-level estimation, all student, school and LA characteristics, explain between 39% and 52.8% of the overall gap for Afrocolombian students in 2010 and 2013, respectively; between 72% and 77.1% of the overall gap for Indigenous students in 2012 and 2010; and between 17.5% and 70.8% of the overall gap for other minority students in 2008 and 2011.

These figures contrast with the estimates using multilevel models, which show that controlling for all these characteristics not always leads to a reduction of the estimated gap. In particular, including the set of all characteristics leads to a 22.3% decrease of the overall gap for Afrocolombian students in 2013 or a 0.7% increase in the gap for this group in 2010; a reduction between 25.2% and 36.7% of the overall gap for Indigenous students in 2008 and 2012, respectively; and a decrease of 20.8% in the overall gap for other minority students in 2011 and an increase of 35.5% of this gap in 2008, as shown in Figure 8.2.

The extent to which LA characteristics are estimated to affect the overall achieve-

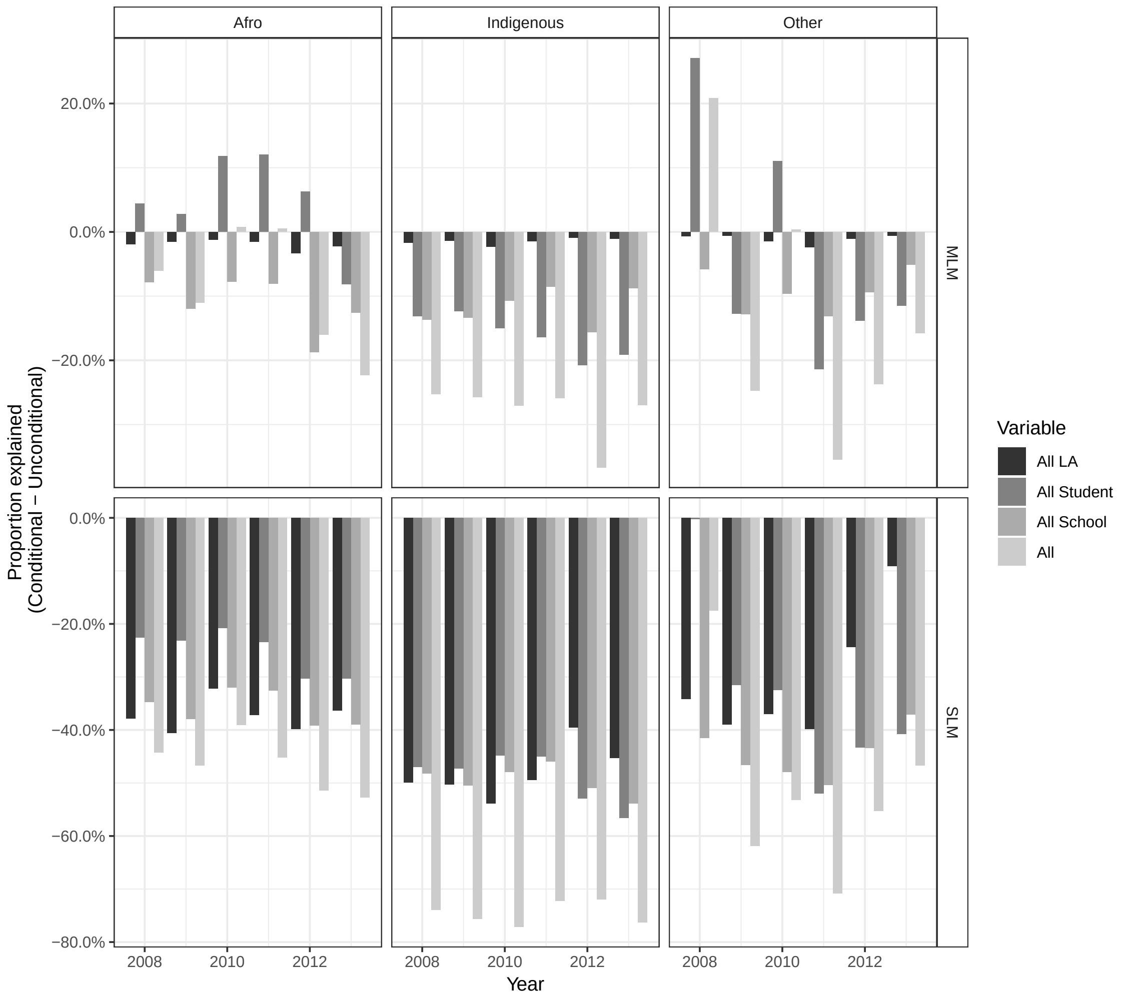

Figure 8.2: Estimated proportion of the overall maths achievement gap that is explained by each set of variables by estimation method

ment gap for Indigenous students is also drastically different when using single-level and multilevel models. While using single-level models leads to the conclusion that LA characteristics explain between 39.6% and 53.9% of the overall achievement gap for Indigenous students (in 2012 and 2010, respectively) , using multilevel modelling leads to the conclusion that LA characteristics only explain between 0.9% and 2.3% of this gap in the same years. These results are the consequence of multilevel estimates that are shrunk towards the within-school achievement gaps. As it becomes clearer with the discussion below, the multilevel estimated changes in the overall achievement gaps mainly reflect the influence of student, school and LA characteristics on the within-school achievement gap.

8.5.1.2 Estimation of the Components of the Gaps

Figure 8.3 shows that, as opposed to the estimation of the overall achievement gap, the estimated unconditional and conditional within-school gap and the school contextual effect of ethnicity are very similar when estimated using a single-level or a multilevel model. Nonetheless, the estimated unconditional LA contextual effect of ethnicity varies depending on the estimation method. For example, the LA contextual effect of Indigenous students is estimated at around 1.18 SD by single-level regression, while the same effect is estimated at 0.52 SD by multilevel regression.

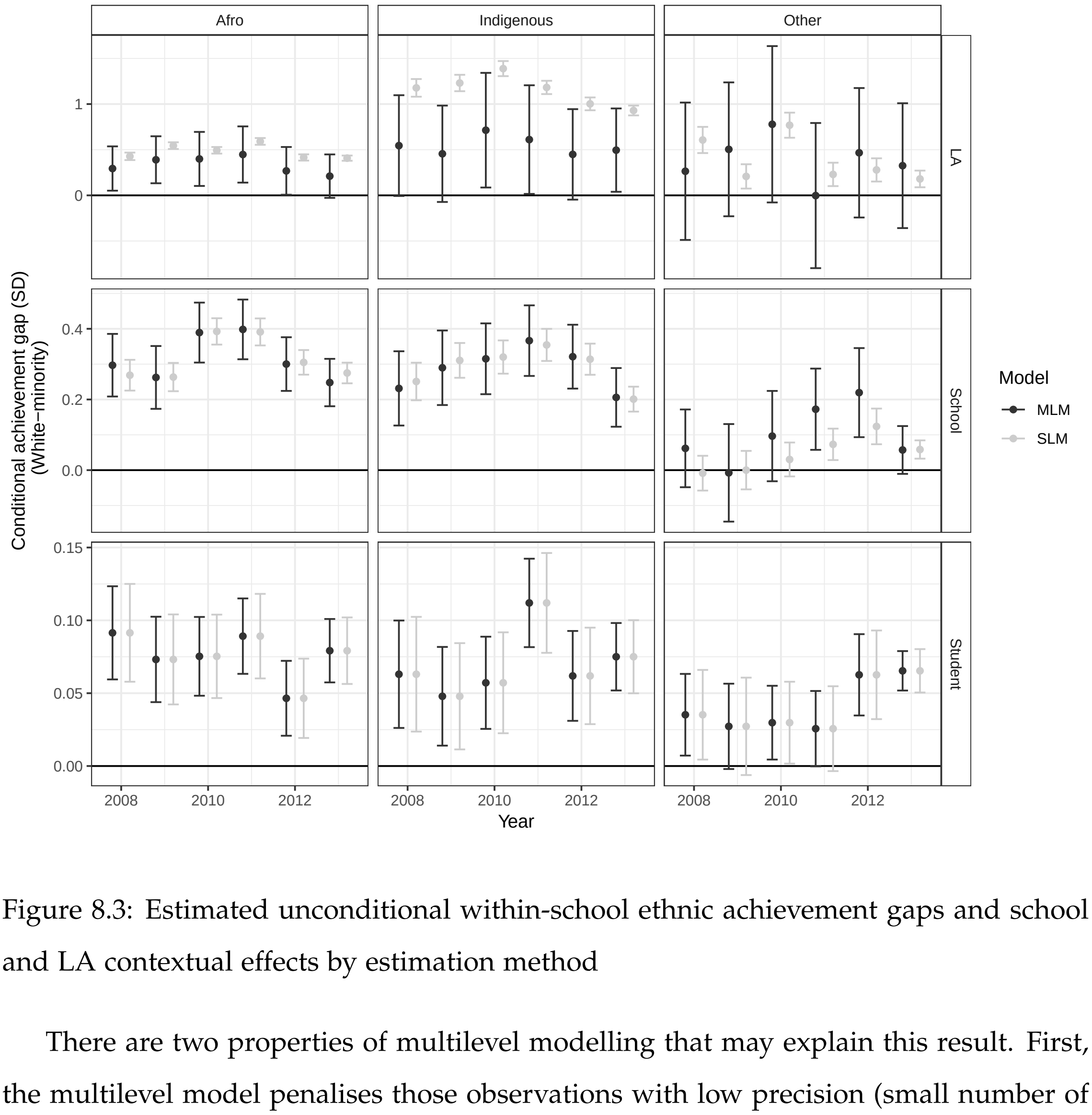

Figure 8.3: Estimated unconditional within-school ethnic achievement gaps and school and LA contextual effects by estimation method

There are two properties of multilevel modelling that may explain this result. First, the multilevel model penalises those observations with low precision (small number of observations or high variability); if there were LAs with few schools or if the schools within some of the LA were highly diverse, the multilevel estimate would assign a lower weight to these observations, and therefore, the multilevel estimates of the LA contextual effects would differ from the single-level estimates. Second, the multilevel estimate of the LA contextual effect may be a combination of effects of a level in-between schools and LA that has not been included in the model (such as neighbourhoods or municipalities) and LA effects; if this were the case, the multilevel estimate would be closer to this intermediate level if they were relatively more homogeneous than the LA. The data provides evidence to support both explanations, as around 34% of the LAs are further divided into municipalities and 5% of the LAs have less than 11 schools.

Figure 8.3 also shows that the multilevel models estimate larger standard errors for the school and LA contextual effects of ethnicity, in comparison to the single-level models, despite using Arellano (1987)'s cluster-robust estimates for the standard errors. This is probably because this kind of robust estimator is less appropriate when the clustering structure is more complex, such as the one in this chapter, which involves a hierarchical structure with three levels (Cameron et al., 2011; Hox, 2010).

Figure 8.4 presents the percentual change in the within-school achievement gaps and school and LA contextual effects after including each set of control variables, for both, single-level and multilevel models. In this case, the effect of shrinkage means that the multilevel estimates of the components of the gap change less when including control variables in the model. For example, both types of models estimate a wider the within-school achievement gap between White and Afrocolombian students and a narrower within-school gap between White and Indigenous students when including all student characteristics. Nonetheless, this increase is estimated at around 17.4% by single-level models and 5.1% by multilevel models in the case of Afrocolombian students, and at around 57.8% by single-level models and 25.1% by multilevel models in the case of Indigenous students.

Similarly, Figure 8.4 shows that the proportion of the school contextual effect that is explained by student and school characteristics tends to be lower when estimated by multilevel than by single-level models. These differences are particularly important in the case of student characteristics. For example, the single-level model estimates that all school characteristics explain around 45.9% of the school contextual effect of Afrocolombian students, while the multilevel model estimates that the same variables explain around 45.7% of this effect.

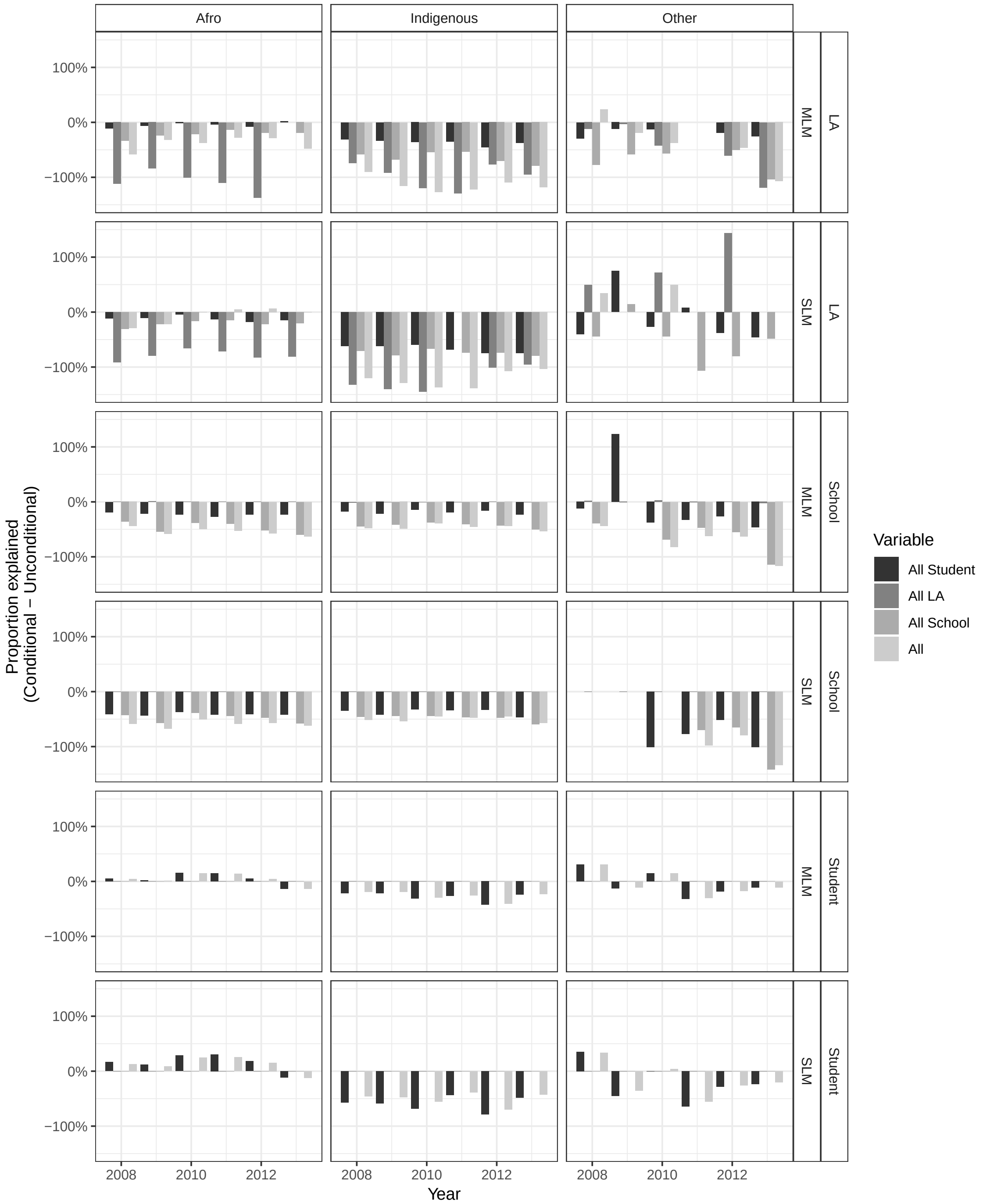

Figure 8.4: Estimated changes in the within-school gap and school and LA contextual effects on maths achievement after controlling for student, school and LA characteristics by estimation method and set of control variables

In turn, single-level models estimate that student characteristics (and therefore, the school composition on these characteristics) explain around 41.4% of the school contextual effect of Afrocolombian students, while multilevel models estimate this figure at around 22.9%. These results are similar for the school contextual effect of Indigenous and other minority students.

In contrast, the estimates for the LA contextual effects are more similar between single-level and multilevel models when conditioning on school characteristics than on student or LA characteristics. Single-level and multilevel models estimate that LA characteristics explain around 80.2% and 110.8% of the Afrocolombian LA contextual effect, and around 136.1% and 93.4% of the Indigenous LA contextual effect, respectively[12]. School characteristics are estimated to explain around 21% and 20.4% of the Afrocolombian LA contextual effect, and around 74.1% and 62.8% of the Indigenous LA contextual effect. Finally, these models estimate that LA characteristics explain around 21% and 20.4% of the Afrocolombian LA contextual effect, and 74.1% and 62.8% of the Indigenous LA contextual effect, according to single-level and multilevel model estimates, respectively. There is no clear pattern for the LA contextual effect of other minority students.

This section showed that both single-level and multilevel models have advantages and disadvantages, and neither of them is flawless for every single research problem. In turn, the choice between them depends not only on the structure of the data that is being used but also on the question that is being addressed. Both kinds of models provide different estimates of the effects of student, school and LA characteristics on the overall achievement gaps and their components. When estimating the overall achievement gap, the difference arises because multilevel models are shrunk towards the within-school gaps, which make single-level models a better choice to analyse the effects of different variables on the overall achievement gap. In turn, when the within-school gaps and the school and LA contextual effects are analysed, the differences in the estimates provided by these methods seem to reflect the influence of particular observations, which have a stronger influence on single-level models than on multilevel models. In this case, multilevel models seem to be a better option.

[12]These changes in the gap above 100% indicate a change of sign in the contextual effect; LAs with a larger proportion of minority students outperform LAs with a larger proportion of White students with the same set of characteristics.

Therefore, in the following analysis and as in chapters 7 and 9, the effect of student, school and LA characteristics on the overall achievement gap will be estimated using single-level models, while the effect of these characteristics on the within-school gap and the school and LA contextual effects will be analysed using multilevel models.

8.5.2 The Role of Student Characteristics

Figure 8.5 shows the single-level estimates of the change on the estimated overall ethnic achievement gap after controlling for student characteristics. These characteristics are represented by different shades and include gender, SES, household size and age. A positive sign of the proportion of the gap that is explained indicates that the overall achievement gap widened after controlling for the variable, while a negative sign implies that the overall achievement gap narrowed when including the variable into the model.

As shown in the figure, for all ethnic groups, SES is the variable that explains the largest proportion of the overall achievement gap, which dominates the joint effect of all the student characteristics. This reflects the fact that gender, age and household size are also correlated with SES. Therefore, when simultaneously including them into the model with SES, the individual effect of these other student characteristics on the achievement gap is not as strong as estimated when including them individually. In particular, Figure 8.5 shows that SES explains between 41.7% (in 2011) and 55.5% (in 2013) of the overall achievement gap between White and Indigenous students, while all student characteristics explain between 44.8% (in 2010) and 56.6% (in 2013) of the gap. Although SES also explains the largest proportion of the overall achievement gap between White and Afrocolombian students, the correlation between SES and other student characteristics is not as strong. In particular, the effect of SES on the overall achievement gap for Afrocolombian students is between 16.2% (in 2010) and 25.3% (in 2013) of the overall achievement gap. The joint effect of all student characteristics on this gap is between 20.8% and 30.3% of the overall gap in 2010 and 2012, respectively.

Figure 8.5 also shows that controlling for gender widens the estimated overall gaps between White and Indigenous students and White and the group of other minority students. Including this variable in the model implies that the ethnic gap is estimated after comparing White girls with minority girls and White boys with minority boys. As girls tend to score lower than boys, if the proportion of girls in the sample were larger for the ethnic minorities than for White students, gender would explain a part of the

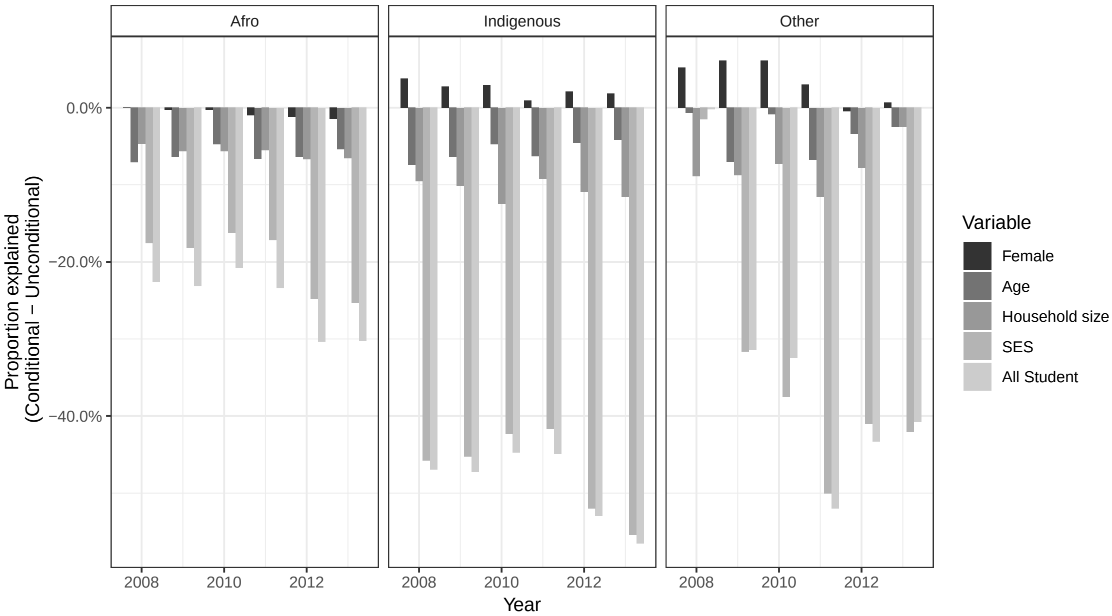

Figure 8.5: Estimated effect of student characteristics on the overall ethnic achievement gaps

overall ethnic achievement gap. This is the case for Afrocolombian students, of which 56.2% are girls, in comparison to 54.8% for White students. Therefore, the estimated gap between White and Afrocolombian students does not change in 2008 and decreases only up to 1.4% (in 2013) when controlling for gender. In turn, the proportion of girls in the sample is lower for Indigenous (51.3%) and other minority students (53.5%). Therefore, controlling for gender incorporates the information that, given the proportion of girls for these ethnic groups, it would be expected that, on average, they scored even lower than their White peers. For that reason, the estimated ethnic achievement gap widens by around 2.4% and 4.1% for Indigenous and other minority students, respectively.

The results are similar when analysing the changes in the gaps after controlling for age and household size. These two variables have a negative correlation with maths achievement and minority students tend to be older and live in bigger households. Therefore, age and household size explain around 6.4% and 5.7% of the overall gap for Afrocolombian, 5.5% and 10.5% for Indigenous and 2.9% and 8.3%for other minority students, respectively, although the results for this last group are less stable over time.

The effect of student characteristics on the overall achievement gap reflects not only the effect of the student characteristics but also the school and LA composition on these characteristics, such as the proportion of girls in the school or the average SES for the school and LAs. To better understand how student characteristics and school and LA composition correlate with the ethnic achievement gaps, it is useful to examine their effects on the within-school gaps and the school and LA contextual effects of ethnicity. These are shown in Figure 8.6.

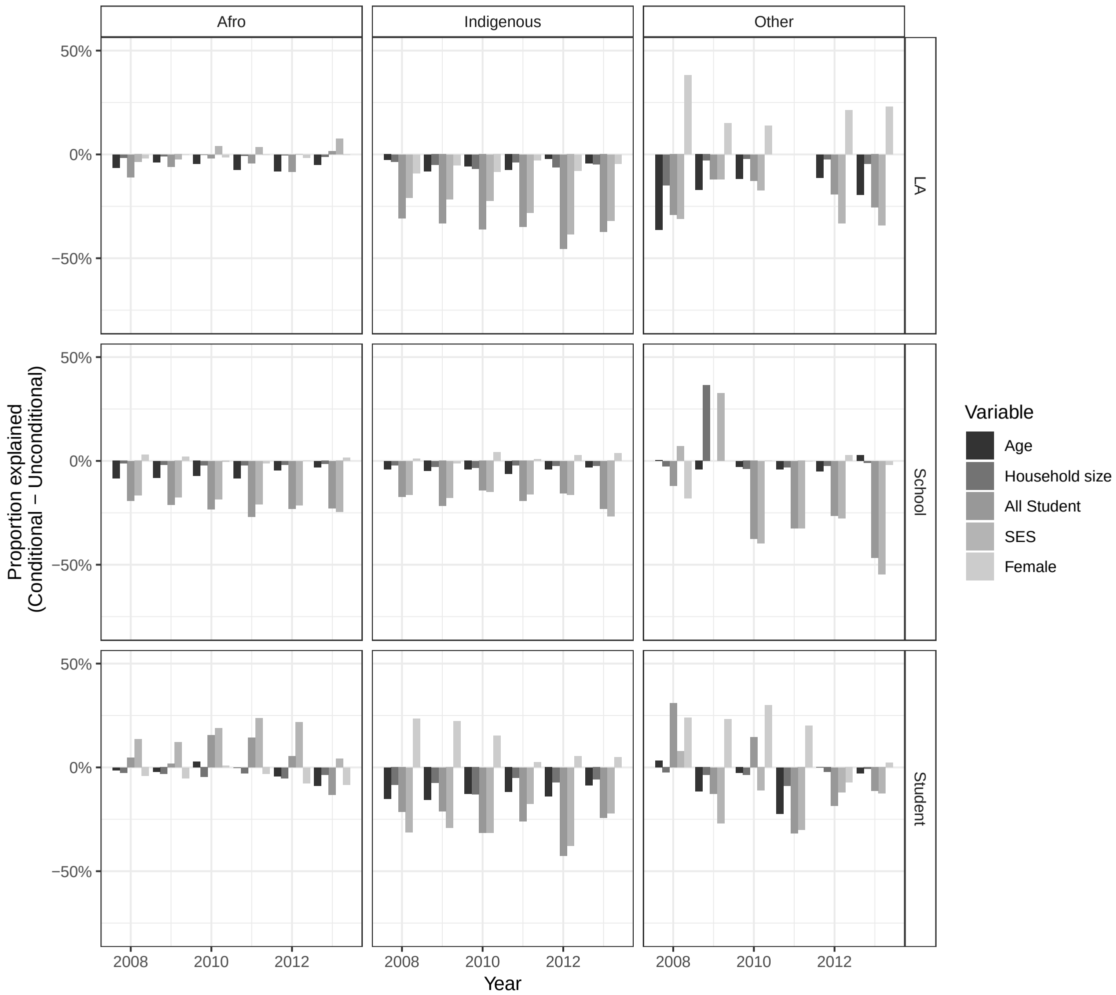

Figure 8.6: Estimated effect of student characteristics on the within-school ethnic achievement gaps and the school and LA contextual effect of ethnicity

This figure shows that SES is associated with the achievement gaps of all minority groups at all levels. Within schools, the estimated achievement gap between White and Afrocolombian students increases between 4.3% in 2012 and 23.6% in 2011, when controlling for SES. This result is a combination of two factors, as shown in the chapter 7. First, Afrocolombian students have an average SES 0.01 SD higher than their White peers attending the same school. In fact, in around 54.6% of schools, the average SES of Afrocolombian students is higher than the SES of their White peers. Second, SES has a positive relationship with maths achievement. What would have been expected is that Afrocolombian students scored higher, given their higher SES. Since this is not the case, the conditional (on SES) within-school achievement gap for Afrocolombian students is wider than the unconditional within school gaps for these ethnic groups. Indigenous and other minority students tend to have a lower SES than their White peers within the same schools and therefore SES explains around 30.2% and 12.3% of the within-school achievement gap, respectively, although the results for the group of other minority students are less stable over time.

At the school level, the effect of SES composition is similar for Afrocolombian and Indigenous students, as this variable explains around 19.7% and 16.3% of the school contextual effect of Afrocolombian and Indigenous students, respectively. In contrast, the LA SES composition is more important for Indigenous than for Afrocolombian students. For Indigenous students, the LA SES composition explains around 25.2% of the LA contextual effect, while for Afrocolombian students, it only explains around 1.8% of this effect. This means that the LA contextual effect of Indigenous students is more strongly correlated with the SES composition than the contextual effect of Afrocolombian students. This may happen because Indigenous students are more concentrated in LAs with a high proportion of low-SES students, while Afrocolombian students may be more dispersed around LAs with different aggregated SES, or because the LA SES composition is more strongly correlated with maths achievement for Indigenous students than for Afrocolombian students.

Figure 8.6 also shows that, in agreement with the results for the overall gaps, the effect of other student characteristics is small, compared to the effect of SES. However, controlling for gender increases the estimated within-school achievement gap between Indigenous and White students between 2.4% in 2011 and 23.4% in 2008, since within schools, the proportion of girls for Indigenous students is larger than the proportion of girls for White students. Besides, controlling for the schools' gender composition also results in an increase of around 1.3% and 2% on the estimated school contextual effect of Afrocolombian and Indigenous students, respectively.

Age and household size also explain a higher proportion of the within-school achievement gap between White and Indigenous students than of the within-school achievement gap between White and Afrocolombian students. Age explains around 13.4% of the within-school gap for Indigenous students, while only around 1.7% of the within-school gap for Afrocolombian students. Similarly, household size explains around 7.3% of the within-school gap for Indigenous students, but only around 3.3% of the within-school gap for Afrocolombian students. In contrast, the school and LA age composition explain around 7.7% and 5.7% of the school and LA contextual effect of Afrocolombian students, and only around 4.1% and 5% of the school and LA contextual effect of Indigenous students. Finally, the school and LA household-size compositions explain a larger proportion of the school and LA contextual effect of Indigenous (2.3% and 4.9%) than for Afrocolombian students (1.8% and 0.7%).

In summary, student characteristics explain a larger proportion of different components of the achievement gap for Indigenous students than for Afrocolombian students. The within-school achievement gap and the school and LA contextual effects of Indigenous students are, to a substantial extent, the reflection of other student characteristics, in particular, SES. For Afrocolombian students, the school contextual effect is also a reflection of the schools' composition SES and other variables. However, the student characteristics explored here do not explain the within-school achievement gap for this ethnic group. Finally, for the group of other minority students, the within-school gap and contextual effects are too close to zero to obtain reliable estimates of the proportion of the gap that these characteristics explain, although these figures also seem to reflect the effects of SES.

8.5.3 The Role of School Characteristics

When examining school characteristics, each particular variable contributes to explaining the overall ethnic achievement gap. Besides, as with the student characteristics, school characteristics explain more of the Indigenous overall student gap than of the overall gap for Afrocolombian and other minority students. Figure 8.7 presents these findings.

As shown in the figure, the most important school characteristic is its type. Whether the school is state or private explains around 17.7%, 25% and 21.8% of the overall achievement gap between White and Afrocolombian, Indigenous and other minority students, respectively. As chapter 4 (section 4.4) showed, students attending private schools tend to obtain higher scores, and a higher proportion of minority students tend to attend state schools, which partially explains the gap. These estimates are relatively low if compared with the combined effect of all school characteristics, which explain around 36.3%, 49.3% and 45% of the overall achievement gap between White and Afrocolombian, Indigenous and other minority students, respectively.

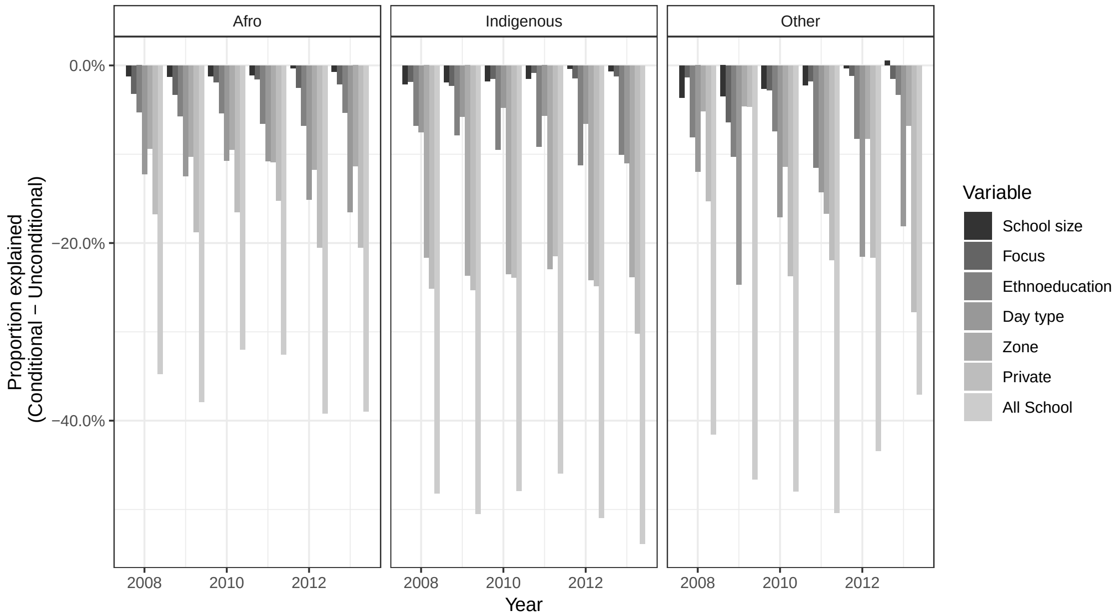

Figure 8.7: Estimated effect of school characteristics on the overall ethnic achievement gaps

Figure 8.7 also shows that whether the school has venues in urban, rural or both kinds of settings is particularly important to explain the overall gap between White and Indigenous students. The effect of the school's zone is around 23.6% of the overall gap between White and Indigenous students, while it is only around 10.6% and 7.5% of the overall gap between White and Afrocolombian and other minority students. In turn, the school-day type (morning, afternoon or full day) explains a larger proportion of the overall ethnic achievement gap for Afrocolombian and the group of other minority students. Including this variable reduces the overall gap between White and Afrocolombian students by 12.4% and the gap between White and the group of other minority students by 17.6%. Whereas the school-day type only explains around 6.1% of the ethnic achievement gap for Indigenous students.

Whether the school offers a curriculum based on ethnoeducation also explains a different proportion of the overall achievement gap for each minority group. For Afrocolombian students, this characteristic explains around 5.5% of the overall ethnic achievement gap. For Indigenous students, the effect of ethnoeducation on the achievement gap varies over time from 6.8% in 2008 to 11.2% of the overall ethnic gap in 2012. The effect of ethnoeducation varies between 3.3% (in 2013) and 11.5% (in 2011) of the overall gap for other minority students, although there is no clear pattern of variation over time.

Finally, school focus (academic, technical, academic and technical or teaching) and school size explain around 2.3%, 1.5% and 1.7%, and 1.2%, 1.7% and 2.4% of the overall ethnic achievement gap between White and Afrocolombian, Indigenous and other minority students, respectively.

These school characteristics explain the overall achievement gaps because of the effect of the school and LA contextual effects of ethnicity. The influence of these characteristics is not necessarily the same for each contextual effect[13], as shown in Figure 8.8.

As the figure shows, for Afrocolombian students, the school contextual effect plays a more important role, while for Indigenous students the LA contextual effect dominates in explaining the overall ethnic achievement gap. For example, the school type (state or private) is the most important variable when explaining the school contextual effect of Afrocolombian students, as it explains between 19.8% and 32.4% of this effect (in 2011 and 2009, respectively), while it only explains up to 11.5% of the LA contextual effect of Afrocolombian students (in 2008). In turn, the school type only explains between 15.4% and 25.7% (in 2011 and 2013, respectively) of the school contextual effect of Indigenous students, while it explains between 35.9% and 48.2% (in 2010 and 2013) of the LA contextual effect of this ethnic group. The school type also explains a larger proportion of the school than of the LA contextual effect of other minority students. Nonetheless, the unconditional contextual effects for this ethnic group are close to zero, making small changes in the conditional contextual effects have a large effect on the estimated proportion of the effect that is explained by any variable. For this reason, Figure 8.8 does not display the estimated changes in the contextual effects for this ethnic group in some of the years in the sample.

Figure 8.8 also shows that the role of school zone (urban, rural or both) is different for the school and LA of Afrocolombian and Indigenous students. Again, for Afrocolombian students the school zone explains a larger proportion of the school contextual effect (11.7%) than of the LA contextual effect (3.6%), which contrasts with the results for In-

[13]Please notice that school variables cannot explain variation within schools. Therefore, the changes in the overall achievement gap after including school-level variables cannot be attributed to changes in the within-school achievement gap.

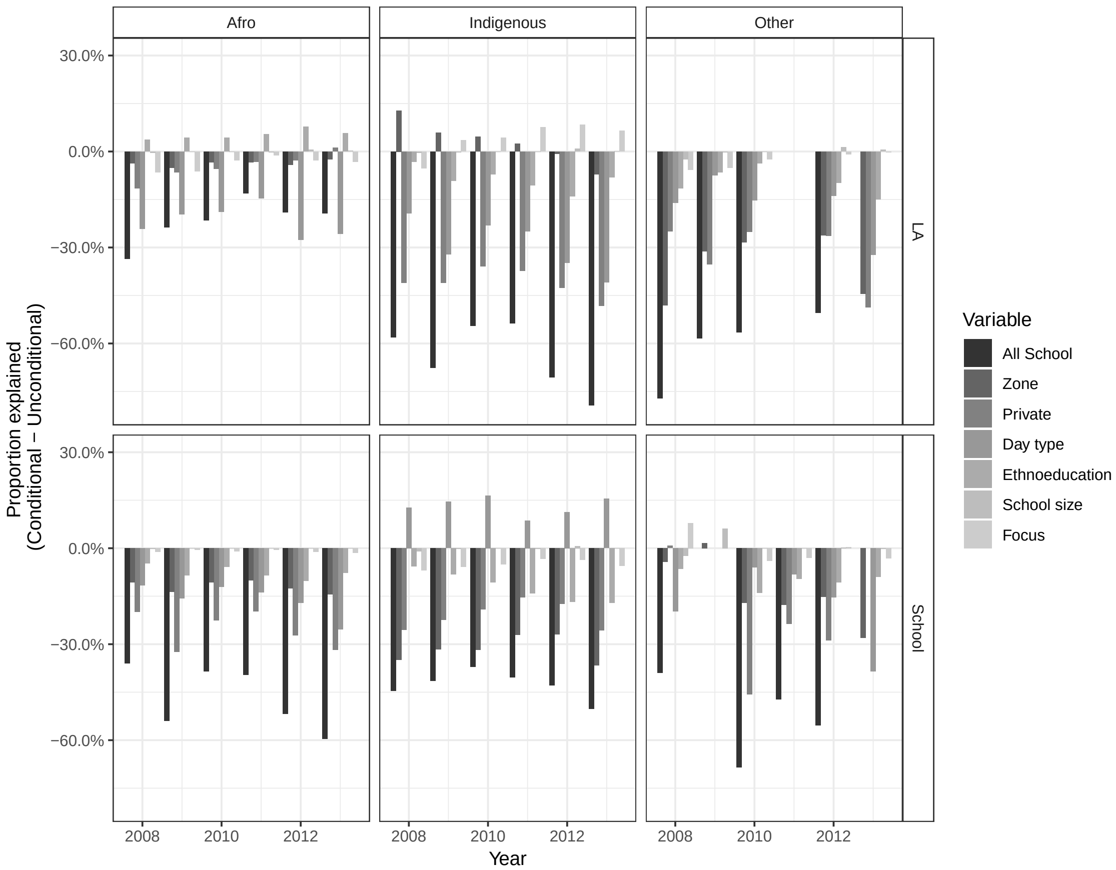

Figure 8.8: Estimated changes in the school and LA contextual effects of ethnicity by school characteristic

digenous students, for who school zone explains around 31.7% of the school contextual effect, but does not explain the LA contextual effect. In turn, when comparing LAs with the same proportion of schools in each zone, the LA contextual effect is around 3.5% larger than when comparing two random LAs. Students in urban schools tend to obtain higher scores than students in rural schools, and Indigenous students are more likely to attend rural schools than their peers from other ethnic groups. LAs with a high proportion of urban schools also tend to have a higher average maths achievement. However, LAs with a high proportion of Indigenous students also have a high proportion of urban schools. For example, Uribia is an LA where around 87.2% of students taking the SABER 11 exam each year are Indigenous students and where around 61.2% of the schools are in an urban area. This contrasts with the average 49.4% of schools in urban areas in LAs where all students who took the SABER 11 exam are White.

The school day type (morning, afternoon or full day) also has a different effect on the school and LA contextual effects of Afrocolombian and Indigenous students. For Afrocolombian students, the school day type usually explains a smaller proportion (around 14.8%) of their school contextual effect than of their LA contextual effect (around 21.9%). In turn, controlling for school day increases the estimated school contextual effect of Indigenous students in around 13.6% and explains between 19.4% and 40.9% of the LA contextual effect of this ethnic group (in 2008 and 2013, respectively). Students attending schools with a full-day program tend to obtain higher scores than students attending schools in the morning, who also tend to score higher than students attending in the afternoon. Minority students, especially Afrocolombian students, are less likely to attend schools offering a full-day program. However, Indigenous students are also less likely to attend schools in the afternoon in comparison to White students, which explains why the estimated school contextual of Indigenous students increases when controlling for school-day type.

During the first two years in the sample, whether the school offered an ethnoeducation program explained a similar proportion of the school contextual effect of Afrocolombian and Indigenous students (around 5% and 8% in 2009, respectively). However, since 2010 ethnoeducation explains a higher proportion of the school contextual effect for Indigenous (between 10.7% and 17.1% in 2010 and 2013) than for Afrocolombian students (around 8.1%). The proportion of schools offering an ethnoeducation program also explains between 8.7% of the LA contextual effect of Indigenous students. In contrast, controlling for ethnoeducation increases the estimated LA contextual effect of Afrocolombian students. LAs with the highest proportions of schools offering ethnoeducation programs also have a high proportion of Indigenous students, while LAs with high proportions of Afrocolombian students have a similar proportion of schools with ethnoeducation programs to LAs with a high proportion of White students.

Finally, while the school focus (academic, technical, academic and technical or teaching) explains a small proportion (around 1.1%) of the school contextual effect of Afrocolombian students, it is relatively more important to explain their LA contextual effect (explaining between 1.2% and 6.4% of this effect in 2011 and 2008, respectively) and the school contextual effect of Indigenous students (explaining around 5.3% of the effect). However, the LA contextual effect of Indigenous students is around 5.4% higher when estimated after comparing LAs with a similar proportion of schools offering these foci. That is, despite the LAs with a high proportion of Indigenous students having a similar offer of schools with academic focus to LAs with a high proportion of White students, students in predominantly Indigenous LAs score lower than students in predominantly White LAs. For none of the minority groups, does school size play an important role in explaining their school and LA contextual effects.

Similarly to what happens with student characteristics, this section shows that the school and LA contextual effects for Afrocolombian and Indigenous students are neither explained in the same way nor by the same set of factors. Instead, each school characteristic contributes differently to explaining the school and LA contextual effects of each ethnic group. The analysis of the group of other minority students is complicated, given the sensitivity of the estimates to the inclusion of different control variables, which are not stable over time. Nonetheless, the importance of school type in explaining the overall gap and the school and LA contextual effects of ethnicity is a consistent finding across ethnic groups.

8.5.4 The Role of Local Authority Characteristics

The LA characteristics that are analysed can be divided into two broad groups: In the first group are the variables that are related to the funding of education, which include fiscal performance and national transfers for investment in infrastructure and educational resources ('Quality transfers') and transfers to cover the running costs ('Running transfers'). In the second group are the variables about violence in the LA, which include the intensity of violence ('Conflict Int' and 'Crime Int') and the pressure exerted by it ('Conflict Pres' and 'Crime Pres'). Figure 8.9 shows how the overall achievement gaps change after including each of the LA characteristics in the model.

As shown in this figure, while variables in the first group explain a larger proportion of the overall gap for Indigenous students, variables in the second group are more important to explain the overall gap for Afrocolombian students. However, transfers for investment in quality and fiscal performance explain an important proportion of the overall achievement gap for all minority groups. Transfers for quality explain around 31.4%, 42% and 33.1% of the overall achievement gap between White and Afrocolombian, Indigenous and other minority students. Similarly, fiscal performance explains around 32% of the overall maths achievement gap between White and Afrocolombian and Indigenous students, and around 24% of the overall gap for other minority students. These two variables are correlated with other LA characteristics, as shown in the data chapter (section 4.5). Thus, combining them does not explain a much larger proportion of the overall gap. Specifically, all the LA characteristics explored in this chapter explain around 37.5%, 49.7% and 35.6% of the overall gap for Afrocolombian, Indigenous and other minority students, respectively.

Transfers to cover the running costs explain a slightly lower proportion of the overall ethnic achievement gaps, as they explain around 25.9%, 30.1% and 28.8% of the overall achievement gap between White and Afrocolombian, Indigenous and other minority students, respectively. Hence, while fiscal performance explains a similar proportion of the achievement gap for Afrocolombian and Indigenous students, both kinds of transfers explain a larger proportion of the overall achievement gap between White and Indigenous students.

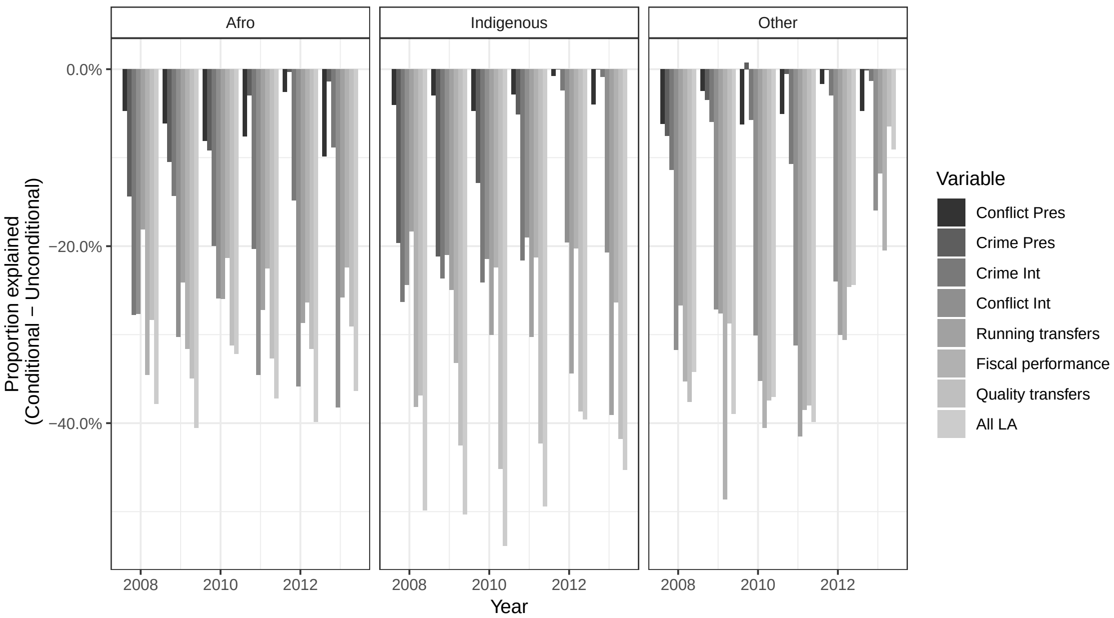

Figure 8.9: Estimated change on the overall achievement gaps by LA characteristic

In terms of the role of conflict, Figure 8.9 shows that the intensity of the conflict explains a larger proportion of the overall ethnic achievement gap for all ethnic groups than the pressure of the conflict. Additionally, while the actions perpetrated by guerrillas explain a larger proportion of the overall achievement gap for Afrocolombian students, the actions from organised crime explain a larger proportion of the overall achievement gap for Indigenous students. The intensity of the conflict ('Conflict Int') explains around 32.4% of the achievement gap for Afrocolombian students, while around 20.8% and 28.6% of the achievement gap for Indigenous and other minority students, respectively. In contrast, the intensity of organised crime ('Crime Int') explains around 22.6% of the achievement gap for Indigenous students, while around 17.4% and 5.9% of the achievement gap for Afrocolombian and other minority students. These figures for intensity contrast with those of pressure, which are around 6.9%, 3.5% and 4.9% of the achievement gap for Afrocolombian Indigenous and other minority students, in the case of conflict ('Conflict Pres'), and 6.1%, 9% and 0.3% in the case of organised crime ('Crime Pres'), respectively.

All these LA characteristics explain the overall achievement gap by changing the LA contextual effect of ethnicity. These changes, which are estimated using multilevel models, are shown in Figure 8.10. Changes in the contextual effect of the group of other minority students are noisy since the estimates of this effect are close to zero. Thus, controlling for LA characteristics produces a large percentage change in the contextual effect, even if the resulting conditional effect is still close to zero.

The figure also shows that these LA characteristics explain a larger proportion of the LA contextual effect of Afrocolombian students, in comparison to the same contextual effect of Indigenous students. In particular, Figure 8.10 shows that fiscal performance explains around 65.6% of the LA contextual effect of Afrocolombian students, but only around 30.3% of the LA contextual effect of Indigenous students. Students in LAs with poor fiscal performance tend to obtain lower test scores and LAs with higher proportions of Afrocolombian and Indigenous students also tend to have poorer fiscal performance, which explains this result.

Given the rules for national transfers presented in the context chapter (section 3.3.2.1), LAs with higher proportions of Afrocolombian and Indigenous students are assigned a larger amount of resources per student. As these transfers are based on needs per student, they are associated with lower achievement, as shown in the data chapter (section 4.5). For this reason, transfers for quality and transfers for running costs explain around 76.4% and 60.6% of the LA contextual effect of Afrocolombian students and around 74.8% and 37.3% of the LA contextual effect of Indigenous students, respectively.

When examining the role of crime and conflict, the results reveal that the actions of both, organised crime and guerrillas explain a larger proportion of the LA contextual effect of Afrocolombian students than of Indigenous students. This way, while crime intensity explains around 51.5% of the LA contextual effect of Afrocolombian students, and 15.6% of the LA contextual effect of Indigenous students, conflict explains around

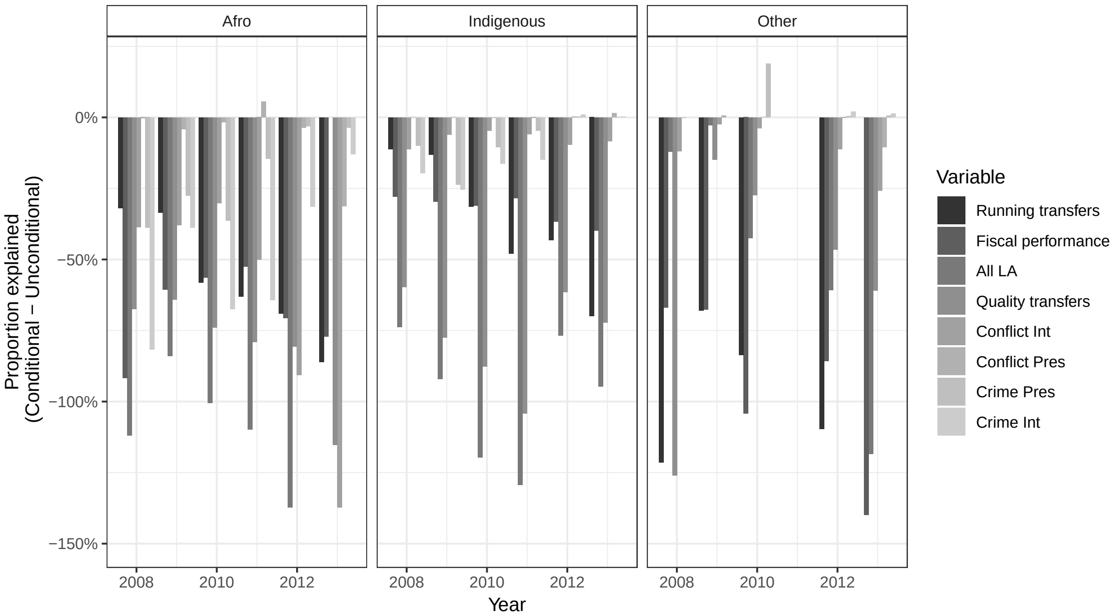

Figure 8.10: Estimated changes in the LA contextual effect of ethnicity by LA characteristic

44.3% and 7.3% of the LA contextual effect of Afrocolombian and Indigenous students, respectively.

These results imply that the LAs with higher proportions of minority students are also the LAs with worse fiscal conditions and that are more severely affected by conflict and crime. However, the ability for these variables to explain the overall achievement gap and the contextual effect of ethnicity is different for each ethnic group. While these characteristics explain a large proportion of the LA contextual effect of all ethnic groups, in the case of Indigenous students, this contextual effect mainly reflects the effect of the composition of LAs in terms of student and school characteristics.

8.5.5 Combined Effect of Student, School and LA Characteristics

To summarise the discussion above, this section presents the combined effect of all student, school and LA characteristics on the overall and within-school ethnic achievement gaps and on the school and LA contextual effect of ethnicity. Following the findings of section 8.5.1, Figure 8.11 shows the proportion of the overall achievement gap that is explained by all these variables, as estimated by single-level models. Additionally, Figure 8.12 shows the proportion of the within-school gap and school and LA contextual effects that are jointly explained by these variables, estimated using multilevel models.

As discussed before, this analysis is necessary because student, school and LA tend to be correlated and therefore the proportion of the gap that is explained jointly by these variables is not the same as the sum of the proportions of the gap explained by each variable.

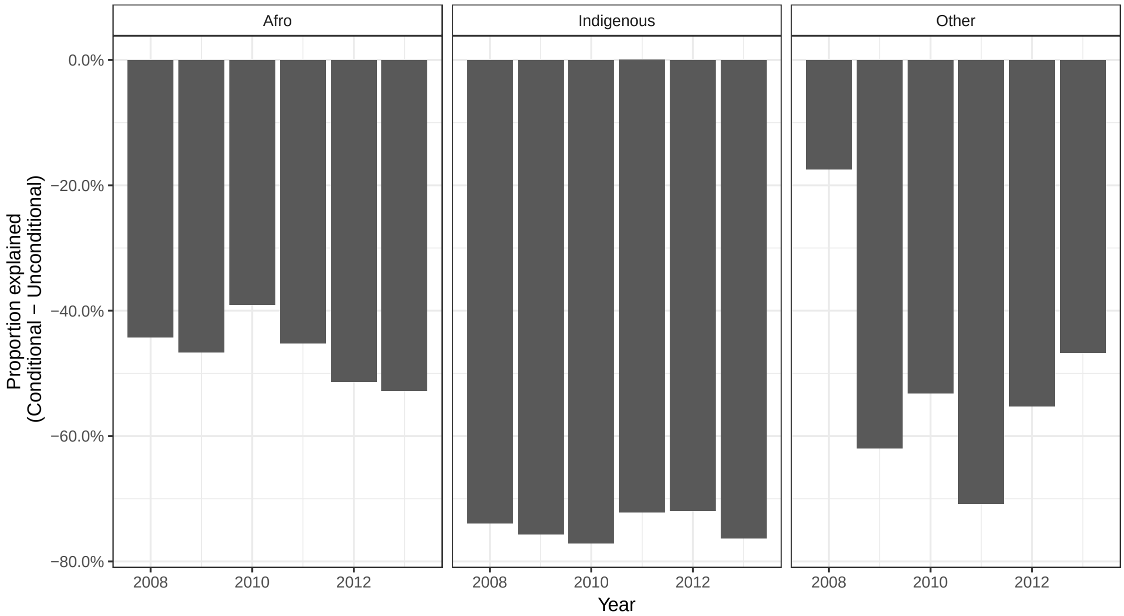

Figure 8.11: Estimated changes in the overall ethnic achievement gap after controlling for all student, school and LA characteristics

Figure 8.11 shows that the proportion of the overall gap that is explained by these characteristics is fairly stable over time for Afrocolombian and Indigenous students, but it is more variable for other minority students. In particular, it is possible to explain around 46% of the gap between White and Afrocolombian students, 75% of the gap between White and Indigenous students, and between 17% (in 2008) and 71% (in 2011) of the overall achievement gap between White and other minority students.

When analysing how student, school and LA characteristics jointly explain the within-school achievement gap and the school and LA contextual effects, the differences among ethnic groups become even more apparent. For Afrocolombian students instead of explaining the within-school achievement gap, considering SES reveals these students score lower than their White peers with the same SES within the same school. Although other student characteristics explain some of the gap, this is not enough to compensate for the effect of SES. In turn, it was possible to explain between 19% and 41% of the within-school gap between White and Indigenous students in 2008 and 2012, respectively. In contrast, the school composition and characteristics explain around 55% of the school contextual effect of Afrocolombian students, but only around 46% of the school contextual effect of Indigenous students. Finally, while only between 28% and 58% of the LA contextual effect of Afrocolombian students (in 2011 and 2008) can be explained by student and school composition and LA characteristics, these variables explain all the contextual effect of Indigenous students and in fact, show that when all these characteristics are the same, students in LAs with a higher proportion of Indigenous students obtain higher maths test scores.

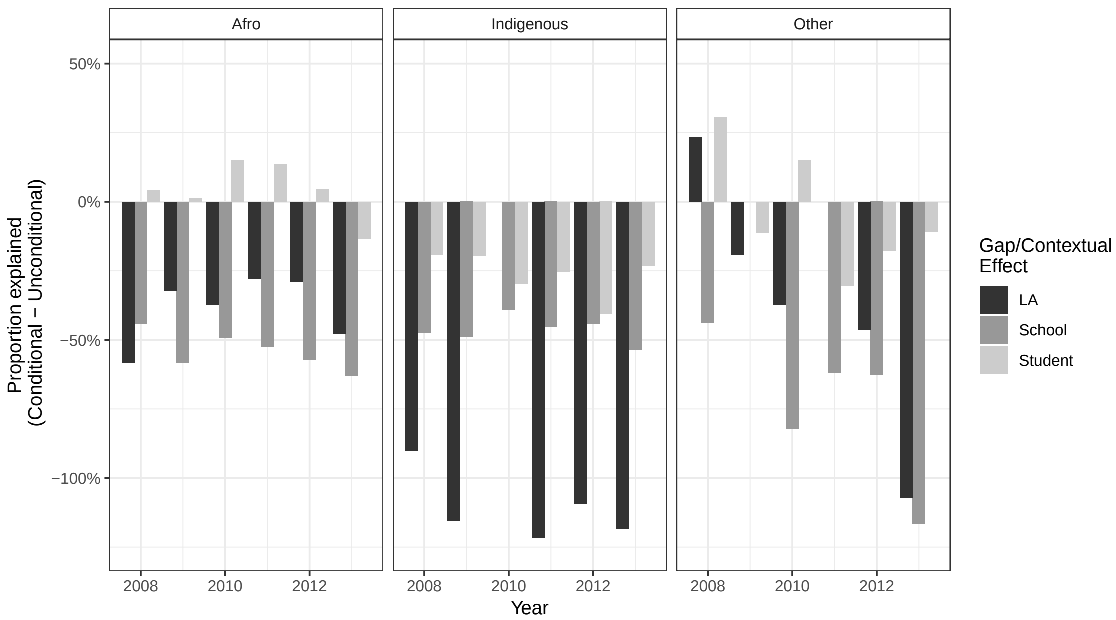

Figure 8.12: Estimated changes in the within-school achievement gap and the school and LA contextual effect after controlling for all student, school and LA characteristics

Examining both the overall ethnic achievement gap and its components highlights the importance of considering the achievement gap as a multilevel phenomenon. Notably, such analysis shows that the fact that student characteristics partially explain the gap is more closely linked processes associated with the school and LA composition, than to student-level processes. The implications of this and other findings are discussed in the next section.

8.6 Discussion

8.6.1 Summary of Findings

8.6.1.1 Methodological Findings

This chapter first discussed the advantages and disadvantages of using OLS single-level and multilevel modelling when studying how different student, school and LA characteristics explain the overall and within-school achievement gaps and the school and LA contextual effects of ethnicity. Section 8.5.1 showed that, in the Colombian case, the multilevel estimates of the overall achievement gap are strongly shrunk towards the within-school achievement gap. This result reflects the nature of the distribution of the Colombian education system, where students tend to attend schools with very similar peers, and there are substantial differences in the characteristics of schools and their student intake. That is, the largest variation in student characteristics and achievement occurs between schools rather than within schools.

In other education systems, such as the UK one, where students within schools are more diverse (with disadvantaged as well as wealthy students within the same schools), and schools are more similar to each other (schools have access to similar resources), the single-level and multilevel estimates will be more similar, as discussed in the methods chapter (section 5.2).

Since the multilevel estimates of the overall gaps in Colombia are shrunk towards the within-school gaps, using multilevel modelling allows studying the within-school gap, but not the overall gap. In other words, using multilevel modelling to study the ethnic achievement gap is not necessarily equivalent to understanding it as a multilevel problem. If a contextual-effect model (or equivalent) is estimated to study the components of the gap, then Myers et al. (2004)' finding that OLS and multilevel regression led to similar results holds. In this case, the advantage of using multilevel modelling is that unreliable observations do not have a substantial effect on the estimation of the components of the gap and how they change when controlling for other variables.

Therefore, how important choosing between multilevel or single-level modelling is depends on the context of the analysis. In this case, single-level models were preferred for analysing the overall achievement gaps, since they do not provide shrunken estimates. This is at the cost of modelling the error structure more appropriately to guarantee accurate inference, given that cluster-robust estimators do not adequately capture the complexity of the error structure (Cameron et al., 2011).

Multilevel analysis was then used to estimate the changes in the within-school gaps and the school and LA contextual effects of analysis, as precision weighted estimates were an advantage to avoid the influence of unusual observations while providing consistent estimates. This decision is in line with Bell et al. (2019)'s advice of specifying a contextual effects model when using multilevel modelling. Nonetheless, these decisions were taken following the research questions that this chapter aimed to address, but they do not imply that all researchers should take the same set of decisions when performing a similar analysis.

In this case, the combination of single-level and multilevel modelling allowed examining the ethnic achievement gap as a multilevel phenomenon, considering both, the overall ethnic achievement gap and its components (the within-school gap and the school and LA contextual effects of ethnicity). Previous research has attempted to simultaneously consider the role of student and school characteristics in explaining the ethnic achievement gap. As discussed in section 8.2, these attempts are based on versions of the Oaxaca (1973) decomposition or comparisons between the overall unconditional achievement gap and estimated achievement gaps after including a set of controls. When these controls include the school ethnic composition (Bali & Alvarez, 2003; Myers et al., 2004; Noe et al., 2005, e.g.,) or school fixed effects (Arteaga & Glewwe, 2019; McEwan, 2004; McEwan & Trowbridge, 2007; Yang et al., 2015, e.g.,), these approaches become problematic for two reasons. First, the change in the coefficient for minority status does reflect not only the role of other student and school characteristics but also the proportion of the gap that is attributed to differences between schools. This is because the inclusion of these variables implies that the estimated parameter no longer estimates the overall ethnic achievement gap but the within-school achievement gap, as explained in chapter 6. Second, they do not allow analysing whether these characteristics are linked to ethnicity throughout student or school-level processes.

These two problems may lead to misleading conclusions when using any of these approaches by attributing the effect of student characteristics uniquely to student-level processes and ignoring the role of school composition. As discussed in the next section, the methodological approach followed in this chapter allows for further insights into why student, school and LA characteristics can explain the ethnic achievement gap.

8.6.1.2 Substantive Findings

This chapter showed that there are differences in the way student, school and LA characteristics contribute to explain the ethnic achievement gap for each ethnic group. Therefore, analysing each group separately improves the understanding of ethnic achievement gaps in Colombia. For example, the characteristics considered in this chapter explain a larger proportion of the overall gap between White and Indigenous students (74.8%) than between White and Afrocolombian (45.9%) and other minority students (54.2%).

The characteristics considered in this chapter were similar, although not the same, like the ones included in Sánchez-Jabba (2011)'s analysis, who was able to explain 60.6% of the overall achievement gap between White and minority students, considered as an aggregate. Sánchez-Jabba (2011) estimate was a weighted average of the ability of these characteristics to explain the overall achievement gaps for each ethnic group, which ignores the particularities of each of them that his chapter has revealed. The inclusion of LA characteristics in this chapter does not explain a much larger proportion of the gap. This is likely to result from a high correlation between the variables considered by Sánchez-Jabba (2011) and the additional variables in this chapter.

Comparisons with international literature are difficult since each study considers a different set of characteristics to explain the gap and focuses on students at various stages in the education system. However, the comparison is still useful to put these results into perspective. As shown in section 8.2, there is a large variation in the ability of student and school characteristics to explain ethnic achievement gaps for different minorities around the world.

In Colombia, SES explains very different proportions of the maths achievement gap between White and Afrocolombian (17.9%), Indigenous (45.5%) and other minority (39.3%) students. In England, SES also explains a dissimilar proportion of the ethnic gap for each minority group. Focusing on Year 9 (age 14) students in 2004, Strand (2011) showed that, for example, SES only explained 25.8% of the gap between White and Black-Caribbean students, but up to 78.4% of the gap for Pakistani students. Strand (2011) also showed that Bangladeshi students with the same SES than White students scored 0.18 SD higher than them, on average. Similarly, in the US, Quinn (2015b) showed that SES explained a different proportion of the achievement gap for Black (74.1%) and Hispanic (67.1%) kindergarteners. Nonetheless, Page et al. (2008)'s finding that parental education explained between 17.2% and 24.2% of the overall White-Black maths achievement gap, is not too different to the finding that SES explains around 17.9% of the overall achievement gap between White and Afrocolombian students.

In Latin-America, the focus has been on the Indigenous-non-Indigenous achievement gap. In Mexico, Blanco (2017) showed that SES explained 48.3% of this gap for 6th grade (age 11/12) students, which is similar to the proportion of the White-Indigenous gap (around 45.5%) that section 8.5.2 showed is explained by SES for 11th grade (age 16/17) students in Colombia. When exploring a further set of student and school characteristics, Blanco (2017) was able to explain 72.4% of the Indigenous-non-Indigenous gap in Mexico, which is also similar to the proportion of the White-Indigenous gap that is explained in this chapter (74.8%). This result is only slightly lower than McEwan (2004)'s finding that family, peer and school characteristics explained 88.4% and 82.4% of the Indigenous-non-Indigenous gap for 6th grade (age 11) students in Bolivia and 8th grade (age 12/13) in Chile, respectively, in 1997. It could be argued that McEwan (2004) were able to explain an additional part of these gaps because they observed the student's language at home, which has been pointed out as an important explanation for the gap, as discussed in the literature review (section 2.4). However, this variable was also included, among other student characteristics, in McEwan and Trowbridge (2007)'s analysis for Guatemala, where they were only able to explain 10.6% of the Indigenous-non-Indigenous gap for 6th grade (age 11/12) students in 2001.

In England, Strand (2011) showed that adding a broader set of student background characteristics, including parental practices and motivation, resulted in a lower proportion of the explained gap (23.3% for Black-Caribbean, 47.1% for Pakistani and a 0.01 SD advantage for Bangladeshi students), given that minority students tend to be more advantaged in these other characteristics, in comparison to White-British students. In general, the evidence for different countries indicates that student characteristics explain either a part or all the achievement gap (Fryer & Levitt, 2004, 2006; Marteleto & Andrade, 2014; Noe et al., 2005). Although in Colombia the overall ethnic achievement gaps for all ethnic groups decrease after controlling for student, school and LA characteristics, this is not always the case when studying the components of the gaps (the within-school gaps and the school and LA contextual effects of ethnicity).

For example, the estimated within-school gap for Afrocolombian students widens when controlling for SES. In other words, Afrocolombian students obtain lower-than-expected scores than their White peers with the same SES within the same school. This happens because Afrocolombian students have a higher average SES than their White peers attending the same schools, but they are more likely to attend schools a socioeconomically disadvantaged composition, as shown in chapter 7 (section 7.4.1). This widening of the within-school gap is compensated in the figure for the overall gap because the school and LA SES composition do explain the contextual effects of Afrocolombian students. Similarly, controlling for school-day type widens the school and LA contextual effects of Indigenous students, because schools and LAs with high proportions of Indigenous students are as likely to offer morning school days as schools with high proportions of White students, but even then obtain lower schools.

The achievement gaps for Afrocolombian and Indigenous students also differed in many other ways, including their relation with ethnoeducation, school type and zone, and LA resources and violence intensity. Most importantly, the LA characteristics rather than LA composition explain most of the overall achievement gap for Afrocolombian students, while the opposite happens with Indigenous students. The multiple differences between Afrocolombian and Indigenous students mainly reflect differences in the circumstances in which these groups live, as described in the data chapter (sections 4.3, 4.4 and 4.5). These differences may arise from two different sources: First, from cultural differences between the ethnicities and second, from differences in the legal treatment that these ethnic groups receive. On the one hand, as two different ethnic groups, Indigenous and Afrocolombian students have different conceptions and values, which also extend to education. On the other hand, as further discussed in section 3.3.1.2, the Colombian legislation also has a different treatment for Indigenous and Afrocolombian groups. These policies are usually linked to the proportion of Indigenous students in a specific territory, which might explain the strong LA contextual effect of Indigenous students. For example, 0.52% of the national resources are transferred to Indigenous territories (Congress of Colombia, 2001), while there is not a similar directive for Afrocolombian students.

Despite their differences, as the LA contextual effect dominates the overall achievement gap for both, Afrocolombian and Indigenous students, explaining this contextual effect contributes to understanding most of the overall achievement gap. This chapter showed that fiscal conditions and the intensity and pressure of violence explain most of the LA contextual effect for these two ethnic groups. Nonetheless, differences in these characteristics are likely to reflect differences in many other features of LAs with different ethnic compositions.

This chapter also showed that considering the different components of the gap when searching for explanations for the ethnic achievement gap provides additional insights as to why these variables explain the gap. In general, student-level characteristics are important explanations for the ethnic achievement gaps not as much as attributes of students, but because of how they are aggregated in schools and LAs with a different ethnic composition. Specifically, student characteristics (SES, household size, age and gender) do not explain the within-school gap between White and Afrocolombian students, and explain around 25.1% of the within-school achievement gap for Indigenous students, respectively. At the same time, the school composition according to these characteristics explains around 22.9% and 18.2% of the negative effect of attending a school with a higher proportion of Afrocolombian and Indigenous students, respectively; and that the LA composition according to these characteristics explains around 5.1% and 35.5% of the negative effect of living in a LA with a higher proportion of Afrocolombian and Indigenous students, respectively.

In the literature, the finding that student characteristics partially explain the achievement gap is used as a call for better understanding student and family processes and interactions and their link to academic achievement, and for policy interventions at the student level, such as differential Free School Meal (FSM) or encouraging better parental practices (e.g. Lenkeit et al., 2015; Sánchez-Jabba, 2011; Stokes et al., 2015; Strand, 2011). Therefore, moving beyond the analysis of the overall gap to consider its components invites to further consider the role of school and LA composition in these characteristics in the formulation of theories and policies around the ethnic achievement gap. Similarly, explanations for the role of school characteristics also need to consider how differences in their availability in LAs with different ethnic composition are linked to the ethnic achievement gap. These policies implications are further discussed in the next section.

8.6.2 Implications for Policy

This chapter showed that the ethnic achievement gaps for all ethnic groups, and Indigenous students, in particular, are largely the reflection of ethnic inequalities in other outcomes. This implies that there is room for policy intervention. Section 8.5.4 showed that LA characteristics explain most of the effect of living in LAs with a higher proportion of minority students (the LA contextual effect of ethnicity) and therefore most of the overall ethnic achievement gaps. Hence, prioritising interventions in LAs with a high proportion of ethnic minorities may be the most effective way of reducing ethnic achievement gaps. Furthermore, the context chapter (section 3.3.2) discussed how LA deliver national policies. The implementation of these national policies could also be revised to secure that they do not disproportionally favour academic achievement in LAs with low proportions of ethnic minorities.

This chapter also showed that the explanations for the ethnic achievement gap are different for each ethnic group. Consequently, existing and future policies at the LA, school and student level should consider tailoring the interventions to fit the needs of each ethnic group and incorporating actions to avoid unintended consequences that broad policy directives may have on each ethnic group.

As further discussed in the next section, the evidence provided in this chapter does not imply any causal mechanisms between the observed variables and the ethnic achievement gaps. Therefore, any policy to tackle them requires further research to understand its potential effects. For this reason, this section only discussed two general guidelines for the aspects that must be considered when formulating any policy of this kind.

8.6.3 Limitations

There are limitations in the methodological and substantive findings of this chapter. Methodologically, the results of the comparison between single-level and multilevel modelling in section 8.5.1 are specific to the Colombian context. As discussed in section 8.6.1.1, the specific characteristics of the context, and in particular, the degree of within- and between-school and -LA clustering of achievement and student and school characteristics. Understanding what specific levels of clustering or other features of the context are more conducive to using single-level or multilevel modelling is out of the scope of this thesis. This could be further explored using Conducting Monte Carlo simulations, which would lead to a more general understanding of when and how single-level and multilevel modelling lead to different results. As explained in section 8.6.4, in the meantime, the recommendation is to be aware of these possible differences and be transparent about the implications of model choice.

Substantively, the results of this chapter do not provide evidence about the causal mechanisms behind the existence of the achievement gaps or the contextual effects or the directions of causality. In turn, the student, school and LA characteristics that have been examined merely show that the estimates of achievement gaps are correlated with different characteristics. Consequently, it should be clear that the estimated changes in the ethnic gaps and contextual effects in this chapter do not imply that any of the variables *causes* the ethnic achievement gap or that ethnicity causes any of the characteristics that have been discussed. For example, the results of this chapter do not allow understanding whether private schools tend to have higher scores because a lower proportion of minority students attend these schools, or if minority students obtain lower scores because they do not attend private schools. These kinds of questions are better solved by accounting for selection in observational data or by taking advantage of a natural experiment. Both approaches are beyond the scope of this thesis.

Additionally, not all possible student, school and LA characteristics can be observed, as the variables are limited by their availability in the database and existing data. Several variables are excluded from the analysis, and it is not possible to understand how they would affect the estimated achievement gaps or the estimated proportions of the gap that are explained by the variables that are included in the analysis. Examples of these variables, which are analysed in the international literature, include the students' prior achievement, motivation and attitude towards maths, time spent doing homework and parental support; the schools' educational resources, physical infrastructure, climate, pedagogical and administrative processes and practices; or the LAs' specific educational policies (Bali & Alvarez, 2003; Becker & Luthar, 2002; Cheng & Starks, 2002; Konstantopoulos, Li, Miller, & van der Ploeg, 2017; Kotok, 2017; Moullin, Waldfogel, & Washbrook, 2018; Strand, 2012). Given the current state of evidence about ethnic achievement gaps in Colombia, it is unknown whether any of these variables would affect the achievement gap.

As chapters 6 and 7, this chapter did not explore interaction effects, assuming that the ethnic achievement gaps and their components are the same regardless of the student, school and LA characteristics. This implies, for example, that the gap between White and Indigenous boys is assumed to be the same as the gap between White and Indigenous girls. This chapter also focused on studying the ethnic achievement gap without considering any relationship between ethnic inequality and academic achievement. These limitations are addressed in chapter 9.

Finally, this chapter shares the limitations from the scope of the thesis, data availability and methodological approach with other research chapters of this thesis, as further discussed in sections 1.4 and 4.6, and 5.4, respectively.

8.6.4 Implications for Future Research

The comparison between single-level and multilevel modelling has two main implications for future research. First, that there is no gold standard when choosing a modelling approach; both methods have advantages and disadvantages and how they affect the analysis will depend on the specific research context and objectives. This chapter illustrated the importance of considering and communicating these decisions. Researchers are therefore recommended to follow a similar approach to be more transparent about the consequences of their methodological choices.

Second, that further research is required to understand how the characteristics of the distribution of the dependent (test scores in this case) and independent variables (e.g., student, school and LA characteristics) result in different estimates for single-level and multilevel models. This could be achieved by conducting Monte Carlo studies that compare the estimates of these models under different degrees of within-school clustering for the dependent variables and types of distributions (including outliers).

The substantive findings of this chapter lead to questions about the causal mechanisms behind these results. This is crucial to understand how any policy would affect ethnic inequality. This chapter showed that the explanations for ethnic achievement gaps are different for each ethnic group, but LAs play an important role. For this reason, there are three evidence needs that could be understood as a priority, given that the existing literature does not provide causal evidence for these links either.

First, it is necessary to understand the differences in the way LAs with differing ethnic composition use the resources that they are allocated for the provision of education. The evidence in this chapter suggests that national transfers for investments in infrastructure and to cover running costs have assigned more resources per/student to LAs in need, as intended by the allocation rules summarised in the context chapter (section 3.3.2.1). Nonetheless, there is no evidence that these resources have a positive impact on achievement. One possibility is that the transfers have a positive effect on the long-term, and therefore, the current achievement is positively associated with past transfers. It is, therefore, necessary to understand how the LAs are using these resources and the effect it has on academic achievement. In the UK, the Education Endowment

Foundation (2019) has developed a toolkit of interventions to improve educational outcomes, according to their cost/effectiveness. This could provide a starting point for understanding how the investments of different LAs result in smaller or larger improvements in academic achievement for their students, although impact evaluations should be conducted for the Colombian context.

Second, it is necessary to understand how and why conflict affects academic achievement and how to mitigate its effects. This chapter showed that the intensity and pressure of conflict partially explain the ethnic achievement gap, especially for Afrocolombian students. The conflict has been shown to have detrimental effects on student enrolment and attendance (Østby & Urdal, 2010; Rodríguez & Sánchez, 2012; Williams, 2004), which could affect academic achievement. Nonetheless, the conflict itself does not necessarily drive the results in this chapter, but other characteristics of the LAs where conflict and crime are prevailing. Examples of these could include a weak institutional framework, lack of access to work, educational and recreational opportunities, lack of positive role models, among others (Arjona, 2016). Understanding which of these factors are linked to the ethnic achievement gap is out of the scope of this thesis, but future research could search for natural experiments or use matching techniques to understand which of these factors cause the ethnic achievement gap.

Third, it is necessary to understand how pedagogical practices and processes differ in schools with varying ethnic composition. Valoyes Chavez (2015) showed that teachers change their pedagogical practices when they teach in schools with a high proportion of Afrocolombian students, adopting strategies that, Valoyes Chavez (2015) argued, hinder the academic achievement of Afrocolombian students. However, in this study it was hard to separate the effects of the school ethnic and SES composition. This kind of study has not been conducted to compare White and Indigenous schools. Understanding how school policy and practice changes in these contexts can aid the design of pedagogical guidelines to tackle the ethnic achievement gap.

8.7 Summary

This chapter built on the findings of chapters 6 and 7 and on Sánchez-Jabba (2011) and the quantitative educational research in Colombia to contribute to the ethnic achievement gap literature by showing that:

1. Shrinkage constitutes both an advantage and a disadvantage of multilevel modelling, depending on the question that is addressed using this method. When analysing the overall ethnic achievement gap in Colombia, multilevel models provide estimates that are shrunk towards the within-school achievement gap, preventing their use for this task. Single-level models are preferred when analysing the overall ethnic achievement gap. In turn, when examining the components of the gap (the within-school gap and school and LA contextual effects of ethnicity), shrinkage is an advantage because it prevents unreliable observations from distorting the results, unlike single-level modelling. The importance of choosing between these two methods is argued to be context-dependent, and researchers are invited to evaluate their advantages and disadvantages for each question and research context.

2. Understanding the ethnic achievement gap as a multilevel phenomenon uncovers that the importance of student characteristics in explaining the gap relies on their distribution among schools and LAs, and not so much on how they affect individual students. For example, SES does not explain the within-school ethnic achievement gap for Afrocolombian students, but the SES composition of schools and LAs partially explain the school and LA contextual effect of ethnicity and therefore the overall ethnic achievement gap.

3. The explanations for the overall ethnic achievement gaps vary by ethnic group, unlike Sánchez-Jabba (2011) assumed. Schools' student intake and characteristics explain a larger part of the overall White-Afrocolombian achievement gap, but SES explains most of the White-Indigenous gap.

9 Are Ethnic Achievement Gaps Uniform Across Subgroups of Students, Schools and Local Authorities?

9.1 Introduction

Chapter 6 showed that the overall ethnic achievement gap (the difference in mean test scores between White and minority students) can be separated into three different components: the within-school achievement gap and the school and local authority (LA) contextual effects of ethnicity. The first component results from comparing the average test scores of White and minority students attending the same school, while the school and LA contextual effects result from comparing the scores of a given student in schools and LAs with a different ethnic composition. Chapters 7 and 8 showed that socio-economic status (SES) and other student, school and LA characteristics partially explain the overall achievement gaps between White and each group of minority students, and the within-school gaps and the school and LA contextual effects of each ethnic group. Nonetheless, differences in student characteristics explain a larger proportion of the achievement gap for Indigenous students than for any other group, while differences in school characteristics are more important to explain the achievement gap for Afrocolombian students.

Until now, it has been assumed that the ethnic achievement gaps and contextual effects are the same for all subgroups of students (e.g. boys and girls, socioeconomically advantaged and disadvantaged), schools (e.g. state and private, urban and rural) and LAs (e.g. well and poorly managed, affected and non-affected by conflict). This chapter relaxes this assumption and investigates how the overall and within-school ethnic achievement gap and the contextual effect of each ethnicity differ between subgroups of students, schools and LAs. This is achieved by including same-level and cross-level interactions and school random effects.

Interaction variables also allow identifying those characteristics that have a different effect on students of each ethnic group. Therefore, the interaction term is interpreted as indicating two different patterns[1]. For example, an interaction between SES and ethnicity shows, first, that the achievement gap is wider at some levels of SES than at others (e.g. the achievement gap may be wider among those with high SES). Second, the interaction term shows that the effect of SES differs among ethnic groups (e.g. an increase in SES has a weaker effect on minority students). The first statement implies that policy interventions aiming to reduce the ethnic gap for students with a specific level of SES should be prioritised and that future research can identify potential lessons from those subgroups with a narrower gap. The second statement implies that any policy that increases SES would disproportionally favour the achievement of one of the ethnic groups, which has the potential to narrow or widen the ethnic achievement gap[2]. Naively assuming that all subgroups of students respond in the same way to a particular intervention may lead to misleading and less effective policies.

The pertinence of future research and policy can also be improved by understanding the level of the education system (student, school or LA) that is linked to these differences in the ethnic achievement gap. For example, if the achievement gap is different in private and state schools, this may be because of differences in their within-school achievement gap or because the school contextual effect of ethnicity is stronger in these kinds of schools. While the first situation would lead to interventions at the student level, the second situation would lead to interventions at the school level.

In most of the achievement gap literature that explores interactions, the concern is usually how the effect of variables that predict achievement is different for ethnic minorities (e.g. Biedinger et al., 2007; Kotok, 2017; Plewis, 2011). This issue is sometimes also examined by estimating separate models for each ethnic group. Examples of this approach include Sánchez-Jabba (2011)'s study in the Colombian context for 2010 and other studies using the Oaxaca (1973)'s decomposition around the world (e.g. Cao & Maloney, 2018; Laborda, Elosúa, & Gómez-Veiga, 2019; Sohn, 2012).

[1]Please refer to appendix A.3.1 for an illustration for these two different interpretations of the interaction term.

[2]This example gives a causal interpretation to the interaction of ethnicity and SES. Nonetheless, SES may reflect the influence of many different variables that are correlated with them, which may include health outcomes or parental involvement, as discussed in section 9.5.3.

Reardon and Galindo (2009) presented a study interpreting interaction terms as differences in the achievement gap. They focused on how the Hispanic-White achievement gaps in maths and reading differed between students with a different immigrant generational status, language use at home and SES. However, Reardon and Galindo (2009) did not consider variation in the gaps between subgroups of schools, random variation in the gap between schools or the different components of the gap.

As further discussed in section 9.2, Reardon and Galindo (2009) is not the only study narrowing their focus to student-level interactions. Indeed, it is unusual for a single study to consider all these different elements simultaneously. Studies including interaction effects normally focus on the overall achievement gap. The peer-effects literature is an area in which contextual effect models that separate the within-school gap from the contextual effect, sometimes including interactions, are often estimated. However, their focus is normally on how different subgroups of students (including students of different ethnicities) are affected by the schools' ethnic composition and not on understanding how the achievement gap changes at different levels of the education system.

Besides this introduction, this chapter includes the following sections:

1. Section 9.2 reviews the ethnic achievement gap literature that considers same-level and cross-level interaction effects and random slopes.

2. Section 9.3 presents the models that will be estimated, the interpretation of their parameters and the data used for their estimation.

3. Section 9.4.1 examines how academic achievement varies among subgroups of students (e.g. boys and girls, students with high and low SES), schools (e.g. state and private) and LAs (e.g. with high and low crime intensity). This analysis provides different achievement scenarios for examining ethnic achievement gaps among these subgroups.

4. Sections 9.4.2 and 9.4.3 examine how the overall ethnic achievement gaps and their components (the within-school ethnic achievement gaps and the school and LA contextual effects of ethnicity) vary among subgroups defined by student, school and LA characteristics.

5. Section 9.4.4 examines if, after considering student, school and LA characteristics, there are between-school differences in the unexplained (residual) within-school

achievement gaps and, if there are, whether these are clustered within LAs.

6. Section 9.5 summarises the findings and discusses their limitations and implications for policy and future research.

7. Section 9.6 summarises the contributions of this chapter to the ethnic achievement gaps literature.

9.2 Literature Review

This chapter estimates models that include interaction terms and random slopes to examine systematic differences in the ethnic achievement gap between groups of students, schools and LAs while examining the different components of the gap (the within-school gap and the school and LA contextual effects of ethnicity). All these elements of analysis are not often combined in a single analysis in the ethnic achievement gaps literature. This section briefly reviews how these elements are considered in the literature and how they relate to the analysis in this chapter.

Stiefel et al. (2007) and Simonite (2005) are particularly close to the type of analysis that this chapter proposes. Stiefel et al. (2007) examined interactions between the gap in academic progress between fifth and eighth-grade between White and Black, Hispanic and Asian students and school characteristics. Using data for New York City in 2000/2001, they found that while a smaller school size and whether teachers received training were associated with a narrower gap for all ethnic groups, a smaller student per teacher ratio and a higher proportion of teachers with master's degree were only associated with narrower gaps for Black and Hispanic students. While this analysis was carried out using a hierarchical linear model, Stiefel et al. (2007) used a fixed-effects model to estimate the within-school gap for the schools with more than five students in each ethnic group. They found that the gaps varied widely between schools, but their approach did not allow them to test if the result held after considering school characteristics.

Simonite (2005) explored fixed and random variation in the gender achievement gap using similar models to the ones used in this chapter. Examining data for UK graduates between 1994/95 and 1999/2000, she estimated both, a conventional and a contextual-effects model including gender random slopes for the probability of achieving first-class honours when graduating in mathematical sciences. Although Simonite (2005) did not discuss the differences in the parameter interpretation of both kind of models, her results indicated that even though there was no overall gender achievement gap, there was a contextual effect of gender (universities with a larger proportion of males were less likely to award firsts). She also found that the positive prior achievement contextual effect is lower for women. As Simonite (2005) discussed, this interaction explained what was initially identified as random variation in the overall gender gap between universities.

9.2.1 Same-level and Cross-level Interaction Effects

Although it is common to estimate the ethnic achievement gap while considering student and school characteristics, it is less common to estimate the gap for specific subgroups of the population. Within the studies that do explore these interaction effects, they usually focus on student-level interactions, such as differences in the ethnic achievement gap by SES, prior achievement or gender. The results are highly dependent on the context. For example, in England and Wales in 1995, Demack et al. (2000) found that at age 16 the achievement gap between White and minority students was wider for children of manual workers than for children of non-manual workers, and that the achievement gap for Black students was wider for boys than for girls, but the opposite was true for the gap for Bangladeshi students. In contrast, Dekkers, Bosker, and Driessen (2000) studied the academic achievement of students of a similar age in the Netherlands in 1995 and found no ethnic achievement gap for girls, although ethnic minority boys scored lower than European students, and a wider ethnic gap for high-SES than for low-SES students.

Reardon and Galindo (2009) showed that there was a 1.1 standard deviations (SD) maths achievement gap between White and first-generation Mexicans, but the gap was of 0.46 SD for third-generation Mexican kindergarteners. They also found that by fifth grade Hispanic students in the bottom 20% of the SES distribution slightly outperformed their White peers, while Hispanic students in the top 20% of the SES distribution scored around a 0.3 SD lower than their White peers. Reardon and Galindo (2009) used data from the US Early Childhood Longitudinal Study following students from kindergarten (age six/seven) in 1998/1999 until grade five (age ten/eleven) in spring 2004. Using the same data, Kieffer (2008) found that the achievement gap between native English speakers and minority language learners was wider in schools with a more favourable SES composition than in socioeconomically deprived schools.

Similarly, using data from a longitudinal study in Flanders, Belgium, Verhaeghe et al. (2018) found that the achievement gap between high-SES Dutch speakers and low-SES non-Dutch speakers varied between 0.91 SD and 1.06 SD at the start of grade one (age 7/8), depending on the language used at home.

Despite the similarity between the approach of these three studies to the interpretation of interaction effects as different gaps between subgroups of students[3], the studies differ from the analysis in this chapter since they did not consider simultaneously random and fixed variation in the gap between schools and they did not explore the different components of the overall achievement gap.

In the case of peer-effect studies, the focus is conventionally on how the school ethnic composition interacts either with the student's ethnicity or with other school-level variables, such as the schools' educational resources and practices. In the first group of studies, Bohrnstedt et al. (2015) estimated a contextual effect model to show that the within-school ethnic achievement gap is the same regardless of the school's ethnic composition, using the National Assessment of Educational Progress (NAEP) data for 8th grade (age 13/14) students in 2011. Bohrnstedt et al. (2015) also found that while there was no Black-White achievement gap for girls, the ethnic achievement gap for boys was 0.5 SD and 0.7 SD, in schools in which Black students accounted for more than 60% and less than 20% of students, respectively. Similar studies focusing on how the contextual effect changes by ethnic group (e.g. Canales & Webb, 2018; Hanushek et al., 2009; Hanushek & Rivkin, 2006; Ready & Reid, 2018) normally show that a higher school proportion of minority students is linked to poorer performance of minority students, but the effect is weaker or non-existing for majority (White) students. Noe et al. (2005) also estimated a contextual-effect model to show that the positive contextual effect of mothers' education was weaker for Indigenous than for non-Indigenous students using SIMCE data for fourth grade (age 9) students in Chile in 1999. They also showed that while non-Indigenous students were negatively affected by larger proportions of Indigenous students, the opposite happened for Indigenous students, who achieved higher test scores.

In the second group of students, Cebolla-Boado and Medina (2011) tested whether the practice of sorting students into classrooms in such a way that low- and high-

[3]As opposed to the interpretation of interaction effects as different effects of predictors for different ethnic groups.

performing students are part of the same classroom was linked to a stronger contextual effect of immigrant primary students in Spain in 2003. However, this hypothesis did not hold after controlling for other student and school characteristics. In Missouri, US, Iv and Hogrebe (2010) found that the school contextual effect of minority students was stronger in schools with higher dropout rates and smaller in schools with more highly qualified teachers, using a sample of 10th grade (age 15/16) students in 2002.

Another approach is separating different regressions for each ethnic group. Bali and Alvarez (2003) used this approach to show that both student and school characteristics had a stronger effect on the maths and reading test scores for Hispanic than for Black students in a school district in California for 2nd to 11th grades (ages 7/8 to 16/17) in 1999/2000. They also estimated a single-level model with interaction effects for student and school characteristics and showed that the Hispanic-White achievement gap for male English-speaking average-SES Free School Meal (FSM)-beneficiary living-with-both-parents fifth-grade students was 33% the size of the Black-White achievement gap for students with the same characteristics.

As summarised in the context chapter (section 3.4.2), in Colombia, Sánchez-Jabba (2011) estimated two separate models for White and minority students to show that, in 2010, students' characteristics that positively affect achievement were more strongly correlated with maths test scores for White than for minority students. For example, Sánchez-Jabba (2011, p.205) estimated that a change in mother's education from no education to postgraduate education was associated with a 15.7% increase in White students' maths test scores, but only with a 12.7% increase in minority students scores. Similar results hold for family income[4]. Besides, Sánchez-Jabba (2011) found that the gender achievement gap was 2.1 percentage points narrower for minority than for White students, favouring boys; the effect of an increase in school fees on maths achievement was 3.4 percentage points higher for minority than for White students; and maths achievement of both White and minority students was similarly negatively affected by larger family size, living in a rural zone, having a job and studying during the night or over the weekends[5]. Sánchez-Jabba (2011)'s approach focused on how student and school

[4] Sánchez-Jabba (2011) did not consider the effects of parental occupation nor of the father's education. He also used school fees as a proxy for school type (state or private). School type is preferred in this analysis.

[5] As described in appendix A.2.1, the data used in this thesis focuses on students in full-time education. Therefore, the sample excludes students attending school at night or over the weekend.

characteristics differentially affected White and minority students and not on how the ethnic achievement gap varied across subgroups of students and schools. Although statistically both approaches are equivalent, the substantive interpretation of the findings in this chapter differs from Sánchez-Jabba (2011), given the aims of this chapter.

None of these studies examined whether the achievement gap varied randomly between schools, that is, whether there was any unexplained variation in the achievement gap between schools. This exclusion does not only imply that between-school unexplained variation is not substantively examined, but the inference for the ethnic achievement gap and these interactions may be over-optimistic (Bell et al., 2019; Nicol, Caplen, Statham, & Browne, 2011). Random between-school variation has been addressed from the perspective of the differential school effectiveness literature, as further discussed in the next section.

9.2.2 Differential School Effectiveness

Within the educational effectiveness literature, there is a relatively less explored branch that focuses on examining if schools are more effective for specific subgroups of students than for others. These studies usually use longitudinal data and focus on student progress, rather than achievement, and the random effects for each school. These random effects represent how much the average progress in test scores in each school differs from the progress that would be expected, given the students' and schools' characteristics. If the progress in a school is higher from that expected, this is interpreted as evidence of school effectiveness. If subgroups of students (e.g. White students) in one school progress more than expected, this is understood as differential effectiveness (Chapman et al., 2016).

Examples of this kind of studies include Nuttall et al. (1989), who estimated a random-slope model for age 16 students in inner London between 1985 and 1987 and concluded that school effectiveness differed by the student's prior achievement. Also using data for age-16 students in inner London, but between 1990 and 1992, S. Thomas et al. (1997) used a multilevel model with random slopes to conclude that there were differential school effects by ethnicity, but schools that were efficient for White students also tended to be efficient for minority students. However, schools with narrower ethnic gaps tended to have wider FSM-status gaps.

These results contrast with Strand (1999, 2010, 2014a, 2014b, 2016), who has consistently found that school effects are the same for students with different prior achievement, gender, ethnicity or FSM status in England and in different LAs in London, at different ages (7, 11, 14 and 16), years (between 1995 and 2008) and subjects. However, Strand's work differs from that of S. Thomas et al. (1997) and Nuttall et al. (1989) in that he has also considered fixed variation in the achievement gap. Specifically, Strand studied fixed and random variation in school effects by gender, ethnicity and FSM status. Strand (1999) studied progress in achievement between age 4 and 7 in a London LA between 1992-1994 and 1995-1997 while including interactions between ethnic groups and FSM status and gender in achievement and progress. According to his findings, students from all minority groups who were entitled to FSM obtained higher scores at age seven than White students who were also entitled to FSM. In other words, within the group of students who were entitled to FSM, the ethnic achievement gap reversed, after controlling for other students' characteristics. The same happened with the gender achievement gap for Caribbean and Black-other students. Similar results were obtained for age 11 students in England in 2004 (Strand, 2010), in LAs in London in 2004, 2005, 2006 and 2008 (Strand, 2014b, 2016) and for age 16 students in England in 2006 (Strand, 2014a). As it is the case with this and other studies including interaction effects (e.g. Collado, Lomos, & Nicaise, 2015; Kotok, 2017; Plewis, 2011), Strand focused on the overall achievement gap, and therefore he did not separate the effect of these variables for students, schools or LAs.

This is also the case for Chatterji (2006), who explained the ethnic, gender and SES achievement gaps in reading for US first graders (age 6/7) in 1999-2001, using the Early Childhood Longitudinal Study. She estimated multilevel models with school random slopes and found that the effects of SES and gender varied between schools, but not the Black-White achievement gap. She also found no interactions between these variables and class and school size, availability of special education, student and parental involvement, teachers' education or the schools' prior achievement and SES composition.

Biedinger et al. (2007) tested whether pre-school attendance accounted for the existence of ethnic achievement gaps in school readiness in a city in Germany. Using data for 6/7-year olds between 2000 and 2005, they estimated a model in which the ethnic achievement gap varied accordingly to the time children spent in pre-school and the school SES and randomly between kindergartens; a model with random slopes and same- and cross-level interactions. They found evidence of random variation in the achievement gap between immigrants and non-immigrants, and that the gap was wider in schools with an advantaged context and for children with less than two years of pre-school exposure, in comparison with children with the same pre-school time but a disadvantaged context.

In Latin America, the study of differential school effects usually focuses on gender, and no consideration has been given to differential ethnic achievement gaps. For example, Muñoz-Chereau (2018) found no substantial evidence of differential school effectiveness by gender for 8th grade (age 13) Chilean students in 2009. Similar findings have been reported for Argentina in 2003 (Cervini, 2006) and 2009 (Formichella & Ibañez Martín, 2014). None of these studies for Latin America considered ethnicity, nor same-level or cross-level interactions between student and school characteristics.

9.2.3 Literature Review Summary

As shown in this section, the international literature provides mixed evidence of different achievement gaps for specific subgroups of students and schools, but also randomly for schools with unobserved characteristics. When the studies include both random and fixed variation between schools, the observed characteristics that interact with the achievement gap seem to explain the variation in the achievement gap between schools fully. The characteristics that are tested for interactions are highly dependent on the study's context and aims, but interactions of the achievement gap with gender and SES are prevalent. This chapter explores if this is the case for ethnic achievement gaps in Colombia while focusing on the different components of the gap (the within-school gap and the school and LA contextual effects of ethnicity). The next section explains how the model that allows this analysis is built.

9.3 Methods and Data

As in chapters 6 and 8, this chapter uses the SABER 11 data for the cohorts between 2008 and 2013, described in chapter 4. Chapter 8 found that student, school and LA characteristics partially explain the overall gap and the within-school gap and the school and LA contextual effects of ethnicity. Chapter 8 also argued that, for the Colombian context, single-level models are more suitable for studying the overall gap and multilevel models for studying its components. This chapter builds on these findings to further examine if the ethnic achievement gap is different for subgroups of students, schools and LAs defined by these characteristics, and explores if there is any remaining unexplained between-school and -LA variation in the achievement gap.

The analysis in this chapter is divided into three parts. The first one explores the overall ethnic achievement gap using single-level regression models with cluster-robust standard errors (Arellano, 1987). The second part explores the different components of the gap (the within-school gap and the school and LA contextual effect of ethnicity) using three-level (students within schools within LAs) random intercept models. The last part considers residual random between-school and LA variation in the within-school ethnic achievement gap by estimating three-level random slope models. In all cases, the analysis implies the construction of complex models with interactions between ethnicity and student, school and LA characteristics.

9.3.1 First Part: Overall Gaps

Hox (2010, p.55-59) and Raudenbush and Bryk (2002) proposed using a 'bottom-up' or 'step-up' approach for model specification. The starting point of this approach is a simple model that only includes main effects for student characteristics, which is then made more complex by adding school and LA characteristics, interactions terms and random slopes. This chapter broadly follows this approach.

The analysis in chapter 8 (section 8.4) provided a list of variables that are correlated with maths test scores and the achievement gap. These variables provide the starting point for this chapter, as they are included in a single-level model with linear effects given by

$$y_{ijk} \quad = \quad \alpha + \beta_A A_{ijk} + \beta_I I_{ijk} + \beta_O O_{ijk} + \mathbf{x}_{ijk}^T \delta_x + \mathbf{z}_{jk}^T \delta_z + \mathbf{w}_k^T \delta_w + e_{y_{ijk}} \tag{9.1}$$

where $e_{y_{ijk}} \overset{i.i.d.}{\sim} N\left(0, \sigma_e^2\right)$[6]. Additionally, $\mathbf{x}_{ijk}$ is a vector of student-level characteristics including SES, gender, household size and age. $\mathbf{z}_{jk}$ is a vector of school characteristics including the school type (state or private), zone (urban, rural or both), day type (full day, morning or afternoon), focus (academic, technical, teaching and academic and technical), whether the school follows a pedagogical approach for ethnic education (ethnoeducation) and the number of students from each school who took the SABER

[6]Please note that the standard errors of the model, which assume homoskedasticity, are replaced with cluster-robust standard errors.

11 exam in each year. Finally, $\mathbf{w}_k$ is a vector of LA characteristics, including national transfers for educational quality and running costs, fiscal performance, and crime and conflict intensity and pressure. These variables are discussed in chapter 3 and described in chapter 4.

As in previous chapters, A_{ijk}, I_{ijk} and O_{ijk} are dummy variables for Afrocolombian, Indigenous and other minority students, respectively. Therefore, β_A, β_I and β_O represent the overall achievement gap between White and Afrocolombian, Indigenous and other minority students, respectively, conditional on other student, school and LA characteristics. As further explained in chapter 6, this model does not separate the effect of ethnicity into its student, school and LA components and therefore the coefficients β_A, β_I and β_O as well as $\boldsymbol{\delta}_{\mathbf{x}}$, $\boldsymbol{\delta}_{\mathbf{z}}$ and $\boldsymbol{\delta}_{\mathbf{w}}$ represent a combination of within- and between-school and LA effects of the student, school and LA characteristics on maths achievement.

Using (9.1) as a starting point, same-level and cross-level interactions are incorporated into the model in three different steps. First, same-level interactions are included if a likelihood-ratio test provides evidence of an improvement of fit when including interaction terms between the dummy variables for all ethnic groups and a student characteristic, in comparison to model (9.1). For example, for the student-level interactions, this implies adding $\left(\theta_A^x A_{ijk} + \theta_I^x I_{ijk} + \theta_O^x O_{ijk}\right) \times x_{ijk}$ to model (9.1) on a separate step for each variable x_{ijk} in $\mathbf{x}_{ijk}$. After testing all potential student-level variables, all those that individually improve the model fit are jointly included in the model. Interactions between the students' ethnicity and school and LA characteristics are included similarly, comparing the improvement of model fit with respect to a model including all interactions with student and school characteristics, respectively. In all cases, the parameter of interest is the conditional overall ethnic achievement gap and not the effect of these sets of characteristics on achievement.

To avoid imposing undesirable restrictions to the estimated parameters, if a variable interacts with the gap for one of the ethnic groups, the model includes interactions for all the ethnic groups, even when that particular interaction is not statistically different from zero for a group. For example, if age interacts with the within-school gap for Afrocolombian but not for Indigenous or other minority students, the model includes interactions for all the ethnic groups. This ensures that the achievement gap is estimated as the difference between Afrocolombian and White students for each age, which is the comparison of interest, as opposed to the difference between Afrocolombian and non-Afrocolombian students. The inclusion of interactions leads to the model

$$\begin{aligned} y_{ijk} = \alpha + \theta_A A_{ijk} + \theta_I I_{ijk} + \theta_O O_{ijk} + \mathbf{x}_{ijk}^T \delta_{\mathbf{x}} + \mathbf{z}_{jk}^T \delta_{\mathbf{z}} + \mathbf{w}_k^T \delta_{\mathbf{w}} + \\ \left(A_{ijk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}_A} + \left(I_{ijk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}_I} + \left(O_{ijk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}_O} + \\ \left(A_{ijk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}_A} + \left(I_{ijk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}_I} + \left(O_{ijk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}_O} + \\ \left(A_{ijk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}_A} + \left(I_{ijk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}_I} + \left(O_{ijk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}_O} + e_{y_{ijk}} \end{aligned} \tag{9.2}$$

In this model, the overall achievement gap β_g, $g = \{A, I, O\}$ is a function of student $\mathbf{x}$, school $\mathbf{z}$ and LA $\mathbf{w}$ characteristics. In other words, the overall achievement gap for each ethnic group g is given by

$$\beta_g = \theta_g + \mathbf{x}_{ijk}^T \boldsymbol{\theta}_{\mathbf{x}_g} + \mathbf{z}_{jk}^T \boldsymbol{\theta}_{\mathbf{z}_g} + \mathbf{w}_k^T \boldsymbol{\theta}_{\mathbf{w}_g} \tag{9.3}$$

which implies that the expected difference in test scores between White and minority students varies between subgroups of students, schools and LAs. For example, the average achievement difference in test scores between White and Afrocolombian students is different between boys and girls, and between state and private schools. This also implies that the coefficient for each ethnic group no longer represents the conditional achievement gap. Instead, it represents the conditional gap for a very particular subgroup of students, schools and LAs: those for which $\mathbf{x}_{ijk} = \mathbf{z}_{jk} = \mathbf{w}_k = 0$. Additionally $\boldsymbol{\theta}_{\mathbf{x}_g}$ ($\boldsymbol{\theta}_{\mathbf{z}_g}$ or $\boldsymbol{\theta}_{\mathbf{w}_g}$) represent how much wider -or narrower- the gaps are for students of a specific subgroup, assuming that they share the same characteristics, excepting for $\mathbf{x}_{ijk}$ ($\mathbf{z}_{jk}$ or $\mathbf{w}_k$).

As in chapter 8, separate models are estimated for each cohort, but the models for all cohorts include the same set of variables $\mathbf{x}$, $\mathbf{z}$, and $\mathbf{w}$. In the unusual case in which the likelihood ratio test led to different conclusions for different cohorts (for example, to include interactions with age), the variable was included if it had a significant effect in at least half of the cohorts. The results of the likelihood ratio tests are shown in appendix A.7.1. An alternative approach would have been to estimate a pooled model using all cohorts simultaneously. The disadvantage of this approach is that it would require increasing the number of estimated parameters to explore whether the interaction terms vary over time. A more parsimonious model was therefore preferred.

9.3.2 Second Part: Components of Gaps

The second part of the analysis extends the models considered so far to consider the different components of the overall gap; the within-school gap and the school and LA contextual effects of ethnicity. This is achieved by including the school and LA proportions of students of each ethnic group into the model; $\overline{A}_{.jk}$, $\overline{I}_{.jk}$ and $\overline{O}_{.jk}$, and $\overline{A}_{..k}$, $\overline{I}_{..k}$ and $\overline{O}_{..k}$, respectively, leading to the contextual effects model

$$\begin{aligned}
y_{ijk} = \alpha + \theta_A^W A_{ijk} + \theta_I^W I_{ijk} + \theta_O^W O_{ijk} + \theta^{CSA}\overline{A}_{.jk} + \theta^{CSI}\overline{I}_{.jk} + \theta^{CSO}\overline{O}_{.jk} + \theta^{CLA}\overline{A}_{..k} + \theta^{CLI}\overline{I}_{..k} + \theta^{CLO}\overline{O}_{..k} + \\
\mathbf{x}_{ijk}^T\boldsymbol{\delta}_{\mathbf{x}} + \mathbf{z}_{jk}^T\boldsymbol{\delta}_{\mathbf{z}} + \mathbf{w}_k^T\boldsymbol{\delta}_{\mathbf{w}} + \\
\left(A_{ijk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}_A}^W + \left(I_{ijk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}_I}^W + \left(O_{ijk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}_O}^W + \\
\left(A_{ijk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}_A}^W + \left(I_{ijk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}_I}^W + \left(O_{ijk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}_O}^W + \\
\left(A_{ijk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}_A}^W + \left(I_{ijk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}_I}^W + \left(O_{ijk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}_O}^W + \\
\left(\overline{A}_{.jk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}}^{CSA} + \left(\overline{I}_{.jk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}}^{CSI} + \left(\overline{O}_{.jk}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}}^{CSO} + \\
\left(\overline{A}_{.jk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}}^{CSA} + \left(\overline{I}_{.jk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}}^{CSI} + \left(\overline{O}_{.jk}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}}^{CSO} + \\
\left(\overline{A}_{.jk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}}^{CSA} + \left(\overline{I}_{.jk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}}^{CSI} + \left(\overline{O}_{.jk}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}}^{CSO} + \\
\left(\overline{A}_{..k}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}}^{CLA} + \left(\overline{I}_{..k}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}}^{CLI} + \left(\overline{O}_{..k}\mathbf{x}_{ijk}^T\right)\boldsymbol{\theta}_{\mathbf{x}}^{CLO} + \\
\left(\overline{A}_{..k}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}}^{CLA} + \left(\overline{I}_{..k}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}}^{CLI} + \left(\overline{O}_{..k}\mathbf{z}_{jk}^T\right)\boldsymbol{\theta}_{\mathbf{z}}^{CLO} + \\
\left(\overline{A}_{..k}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}}^{CLA} + \left(\overline{I}_{..k}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}}^{CLI} + \left(\overline{O}_{..k}\mathbf{w}_k^T\right)\boldsymbol{\theta}_{\mathbf{w}}^{CLO} + e_{y_{ijk}} \quad (9.4)
\end{aligned}$$

where $e_{y_{ijk}} = v_k + u_{1_{jk}} + r_{ijk}$, $v_k \overset{i.i.d.}{\sim} N\left(0, \sigma_v^2\right)$, $u_{jk} \overset{i.i.d.}{\sim} N\left(0, \sigma_{u_1}^2\right)$ and $r_{ijk} \overset{i.i.d.}{\sim} N\left(0, \sigma_r^2\right)$. Which implies that (9.4) is a random-intercept model in which students are nested within schools, which are nested within LAs, and that the conditional ethnic achievement gap has been separated into its conditional within-school gap (β_A^W, β_I^W and β_O^W) and school (β^{CSA}, β^{CSI} and β^{CSO}) and LA (β^{CLA}, β^{CLI} and β^{CLO}) contextual effects[7]. In contrast, the effects of the student, school and LA characteristics have not been separated and therefore the coefficients δ_x, δ_z and δ_w represent a combination of within- and

[7] As a reminder, the within-school achievement gap result from comparing the scores of White and minority students within the same school. The school contextual effect of ethnicity compares the test scores of students in schools that only serve White students with the scores of students in schools that only serve students of each specific minority group, after discounting the effect of the student's own ethnicity. Similarly, the LA contextual effect of each ethnic group compares LA where only White students live with LA where only students of the specific ethnic group live.

between-school and LA effects of the student, school and LA characteristics on maths achievement. The rationale behind this decision is to keep the model as parsimonious as possible to avoid overfitting. Notice that, unlike model (9.2) for the overall gap, model (9.4) allows for the inclusion of interaction effects between each ethnic group and the school and LA contextual effect of ethnicity, testing if the contextual effect of ethnicity varies by ethnic group.

9.3.3 Third Part: Random Slopes

Finally, school-level random slopes are considered for the within-school achievement gap and LA-level random slopes are considered for the within-school achievement gap and the school contextual effect of ethnicity. The results show no evidence of random slopes at the LA level for either the within-school gap or the school contextual effect of ethnicity. In consequence, the structure of the random part of model (9.4) is given by

$$e_{y_{ijk}} = v_k + u_{1_{jk}} + u_{A_{jk}} A_{ijk} + u_{I_{jk}} I_{ijk} + u_{O_{jk}} O_{ijk} + r_{ijk} \tag{9.5}$$

where $r_{ijk} \overset{i.i.d.}{\sim} N\left(0, \sigma_r^2\right)$, $v_{ijk} \overset{i.i.d.}{\sim} N\left(0, \sigma_v^2\right)$ and

$$\begin{bmatrix} u_{1_{jk}} \\ u_{A_{jk}} \\ u_{I_{jk}} \\ u_{O_{jk}} \end{bmatrix} \sim N\left(\mathbf{0}, \Omega_u\right), \Omega_u = \begin{bmatrix} \sigma^2_{u_1} & & & \\ \sigma_{u_{A1}} & \sigma^2_{u_A} & & \\ \sigma_{u_{I1}} & \sigma_{u_{IA}} & \sigma^2_{u_I} & \\ \sigma_{u_{O1}} & \sigma_{u_{OA}} & \sigma_{u_{OI}} & \sigma^2_{u_O} \end{bmatrix} \tag{9.6}$$

This implies that the within-school achievement gap β_g^W for each ethnic group $g = \{A, I, O\}$ is given by

$$\beta_g^W = \theta_g^W + \mathbf{x}_{ijk}^T \boldsymbol{\theta}_{\mathbf{x}_g}^W + \mathbf{z}_{jk}^T \boldsymbol{\theta}_{\mathbf{z}_g}^W + \mathbf{w}_k^T \boldsymbol{\theta}_{\mathbf{w}_g}^W + u_{g_{jk}} \tag{9.7}$$

which implies that the within-school gap not only varies according to the student, school and LA characteristics, but also randomly between schools.

The elements on the diagonal of the matrix Ω_u are the variances of the school random intercept $u_{1_{jk}}$ and random slopes $u_{A_{jk}}$, $u_{I_{jk}}$ and $u_{O_{jk}}$. If these elements are different from zero, there is evidence that the unexplained part of the mean test score for White students (in the case of the random intercept $u_{1_{jk}}$) and the within-school ethnic achievement gaps (in the case of the random slopes $u_{g_{jk}}$) vary randomly between schools. The elements off the diagonal of the matrix Ω_u are the covariances between the random slopes and the random slopes and intercept. For example, $\sigma_{u_{A1}}$ is the covariance between the random intercept and the random within-school gap between White and Afrocolombian students. The interpretation of $\sigma_{u_{A1}}$ depends on the sign of the within-school gap. If White students outperform minority students, a positive value of $\sigma_{u_{A1}}$ would mean that schools with a high unexplained maths achievement for White students also tend to have large unexplained differences in achievement in favour of Afrocolombian students. In contrast, a negative value of $\sigma_{u_{A1}}$ would mean that schools with a high unexplained maths achievement for White students also tend to have large unexplained differences in achievement in favour of White students.

In the differential school effectiveness literature (e.g. Muñoz-Chereau, 2018; Strand, 2010, 2016), the random part of the model is often specified as

$$e_{y_{ijk}} = v_k + u_{W_{jk}} W_{ijk} + u_{A_{jk}} A_{ijk} + u_{I_{jk}} I_{ijk} + u_{O_{jk}} O_{ijk} + r_{ijk} \tag{9.8}$$

where instead of including a school random intercept and three random slopes, four random slopes are included. These random slopes represent how much higher (or lower) than expected students of each ethnic group (White, Afrocolombian, Indigenous and other minority students) score in each school j with respect to the average achievement of each ethnic group. Therefore, this specification is particularly tailored for studying differences in school effectiveness among ethnic groups. In this chapter, the focus is on the between-school variation of achievement gaps between White and minority students; not on the average achievement of each ethnic group. Consequently, the specification of the random part in (9.5) is preferred.

9.3.4 Methods Summary

This chapter uses a bottom-up approach to construct models that allow exploring how the overall achievement gap and its components vary between subgroups of students, schools and LAs. First, this variation is assumed to be associated with observed characteristics using models with fixed slopes. Then any residual differences are allowed to vary randomly between schools using models with random slopes.

9.4 Results

This chapter aims to determine if the achievement gap and its components are different for subgroups of students, schools and LAs. However, focusing only on the ethnic achievement gap presents only a partial picture of the inequality problem, as it focuses on differences between White and minority students and ignores the differences between students of the same ethnic group. Consequently, this section starts with a brief reminder of the importance of differences between students of the same ethnic group, by presenting the maths achievement of different subgroups of students in schools and LAs with different characteristics. Sections 9.4.2 and 9.4.3 then describe how the overall ethnic achievement gap and its components (the within-school gap and the school and LA contextual effects of ethnicity) differ between clearly defined subgroups of students, schools and LAs. Finally, section 9.4.4 presents how the between-school residual variation in the White students' achievement relates to the within-school achievement gaps. The discussion of the results focuses on the findings for 2008, but these hold for other years and are presented in appendix A.7.3.

9.4.1 Differences in Achievement Among Subgroups of Students, Schools and Local Authorities

As explained in section 9.3, the starting point for constructing the model for the study of the ethnic achievement gap for different subgroups of students is the list of variables that explain the ethnic achievement gap, provided by chapter 8 (section 8.4). As discussed in section 8.5, some of these variables are highly correlated, and thus the inclusion of one of them accounts for the information provided by the other variables. For example, LAs with better fiscal performance also tend to have lower crime and conflict intensity and pressure. Different specifications were compared using χ^2 tests[8]. The estimation results of this model with linear effects is presented in Figure 9.1.

[8]The results of these tests to choose the variables to include in the model are presented in appendix A.7.1. The variables that are excluded from this model are the school size and LA's fiscal performance, transfers to cover running costs, and crime and conflict pressure.

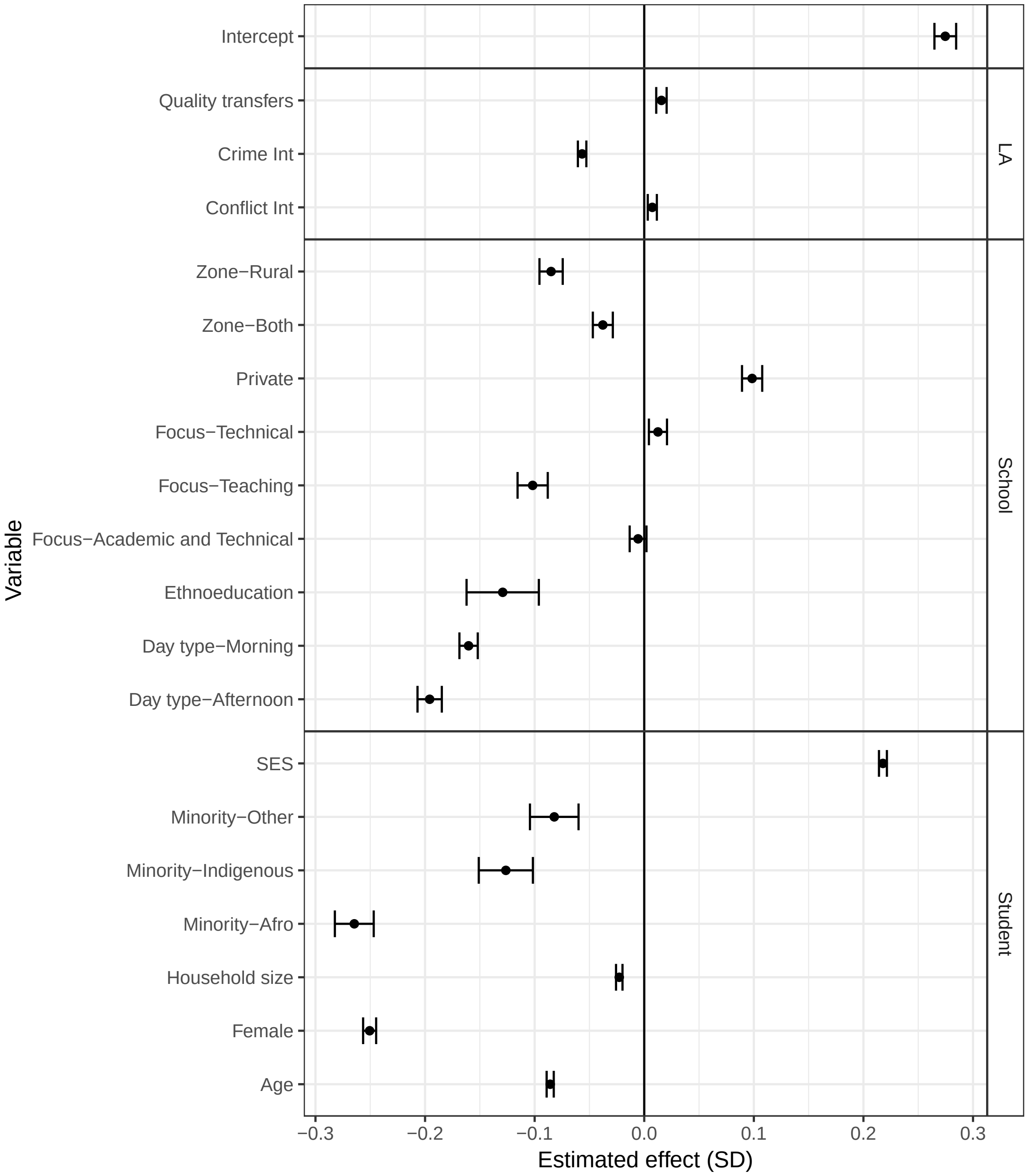

Figure 9.1: Estimated parameters and 95% confidence intervals for the single-level linear effects model (9.1) for maths achievement in 2008

In Figure 9.1, the confidence interval lies to the right of zero for variables with a positive relationship with maths test scores, and to the left of zero for the variables with a negative association with achievement. The effect of those variables whose confidence interval crosses the zero-line is not statistically different from zero with a 95% confidence. As shown in the figure, after considering all other variables, the student's SES and the LA's quality transfers[9] are positively associated with maths achievement, while the student's age and the LA's crime intensity are negatively related to maths test scores.

Additionally, minority students have lower average test scores than White students; girls have lower average maths achievement than boys; students attending school during the morning and afternoon have lower mean achievement than students attending school all day long; the average achievement of students in a school offering a program for ethnic minorities (ethnoeducation) is lower in comparison to the mean achievement of students in schools that do not offer this type of program; students in rural or 'urban and rural' schools have lower average maths test scores than students in urban schools; students in technical schools have higher maths test scores than students in academic schools, but students in schools with a teaching-focus score lower than students in schools with an academic focus, while students in schools with both an academic and technical focus do not have a statistically different average maths score than those students in schools with an academic focus; and students in private schools score higher, on average, than students in state schools. These results are consistent with the findings in the Colombian literature, summarised in the context chapter (section 3.4.1).

The high intercept of this model represents the average test score for White boys with an average age, household size and SES who attend urban state schools offering full-day academic programs that do not involve ethnoeducation in LAs with average quality transfers and crime and conflict intensity. This combination of characteristics, therefore, describe the most advantaged students in the sample.

As shown in appendix A.7.3, the results of the model estimates are relatively stable over time, with three exceptions. First, the association between maths scores and school focus changes over time, which might be related to national policies that strengthen the importance of education for the labour market over academic-oriented education (Atehortúa Cruz, 2006). While students in technical schools tend to score higher than

[9]As described in section 3.3.2.1, quality transfers are those that the central government makes to cover for administrative fees such as printing scorecards and student identifications.

(or, in 2009, the same as) students in academic schools, between 2010 and 2013 students in teaching schools and schools with academic and technical focus obtain higher average maths scores than students in academic schools. Second, the effect of conflict intensity is positive and statistically significant between 2008 and 2012 but is not statistically different from zero in 2013. Third, the effect of the LA quality transfers is positive in 2008 but negative for all other years. This is likely to reflect the changes in the resource allocation rules discussed in section 3.3.2.1, which are more strongly correlated with unfavourable LA characteristics after this year. Since the results are more stable over time for all other variables, the remainder of this section focuses on the findings for 2008 only, the same year that is analysed in chapter 7.

As mentioned in section 9.3, the model presented in Figure 9.1 assumes that the overall ethnic achievement gaps, estimated in 2008 as 0.26 SD, 0.13 SD and 0.08 SD for Afrocolombian, Indigenous and other minority students, respectively, is the same for all students[10], independently of their SES, gender or school type, for example.

To study if this assumption is correct, model (9.2) allows estimating a different overall ethnic achievement gap for different subgroups of students. The estimated coefficients for the full model are presented in appendix A.7.2. According to this model, the overall achievement gap interacts with age, gender, SES, school-day type (morning, afternoon or full day) and school type (state or private). Using this model, it is possible to predict the expected achievement of different subgroups of students. Since the model includes 17 linear predictor variables and 21 interaction terms, constructing different scenarios simplifies the analysis.

Two different school scenarios are considered through this section. The 'favourable' scenario considers private schools with a full school day. The 'unfavourable' scenario considers state schools with a morning school day. In both cases schools are assumed to be in a rural area, have an academic focus and a school program that is not based on ethnoeducation. These scenarios are intended to compare two contrasting cases of achievement while allowing for a sufficient number of observations for each scenario[11].

[10] Equivalently, model (9.1) assumes that the effects of all student, school and LA characteristics are the same for students of all ethnic groups.

[11] In 2008 around 24.2% of students attended private schools and 55.1% of these students have classes all day, while 71.4% of students in state schools only have classes in the morning. The combination proposed by these scenarios implies that there are 37,992 students in the 'favourable' scenario and 77,592 students in the 'unfavourable' scenario. This guarantees that there are at least 20 students of each ethnic group for

Figure 9.2 shows the expected achievement for boys and girls of each ethnic group on each school scenario according to their SES, together with boxplots representing the distribution of the SES for students of each ethnic group, gender and scenario. Other student and LA characteristics are held at their means to facilitate the analysis. As shown in the figure, the SES distribution of students in each school scenario is very different, with high-SES students attending schools in the 'favourable' scenario and low-SES students in the 'unfavourable' scenario. For example, while 71.4% of Afrocolombian, 83.8% of Indigenous and 83.4% of White girls in the 'favourable' school scenario have a SES above the national average, this figure is only of 32.7%, 25.9% and 34.5% for Afrocolombian, Indigenous and White girls in the 'unfavourable' scenario, respectively. The SES distribution is very similar for boys and girls attending schools in the same scenario, especially for boys and girls in the 'unfavourable' school scenario. In the 'favourable' school scenario Afrocolombian girls tend to have a slightly lower SES than Afrocolombian boys, while Indigenous girls tend to have a slightly higher SES than their male peers.

In terms of predicted achievement, Figure 9.2 shows that students with a higher SES tend to obtain higher test scores, which, given the SES distribution for each school scenario, implies that students in the 'favourable' school scenario tend to score above the average (excepting for Afrocolombian students) while students in the 'unfavourable' school scenario tend to score below average. Given that girls also tend to obtain lower maths scores, there are large differences in the maths achievement between girls in the 'unfavourable' school scenario and boys in the 'favourable' school scenario. For example, White, Afrocolombian, Indigenous and other minority girls with an average SES (an unusually high SES for this scenario) in the 'unfavourable' school scenario are expected to score 0.14 SD, 0.37 SD, 0.25 SD and 0.23 SD below the national mean. In contrast, boys with an average SES (an unusually low SES for this scenario) in the 'favourable' school scenario score on average 0.17 SD below the mean in the case of Afrocolombian students, and 0.37 SD, 0.15 SD and 0.32 SD above the mean in the case of White, Indigenous and other minority boys. These results show that there are important differences in achievement between subgroups of students of the same ethnic group.

Figure 9.2 also shows that the differences in the predicted maths test scores between ethnic groups also differ within these subgroups of students. In particular, the test scores

each combination of school scenario and gender.

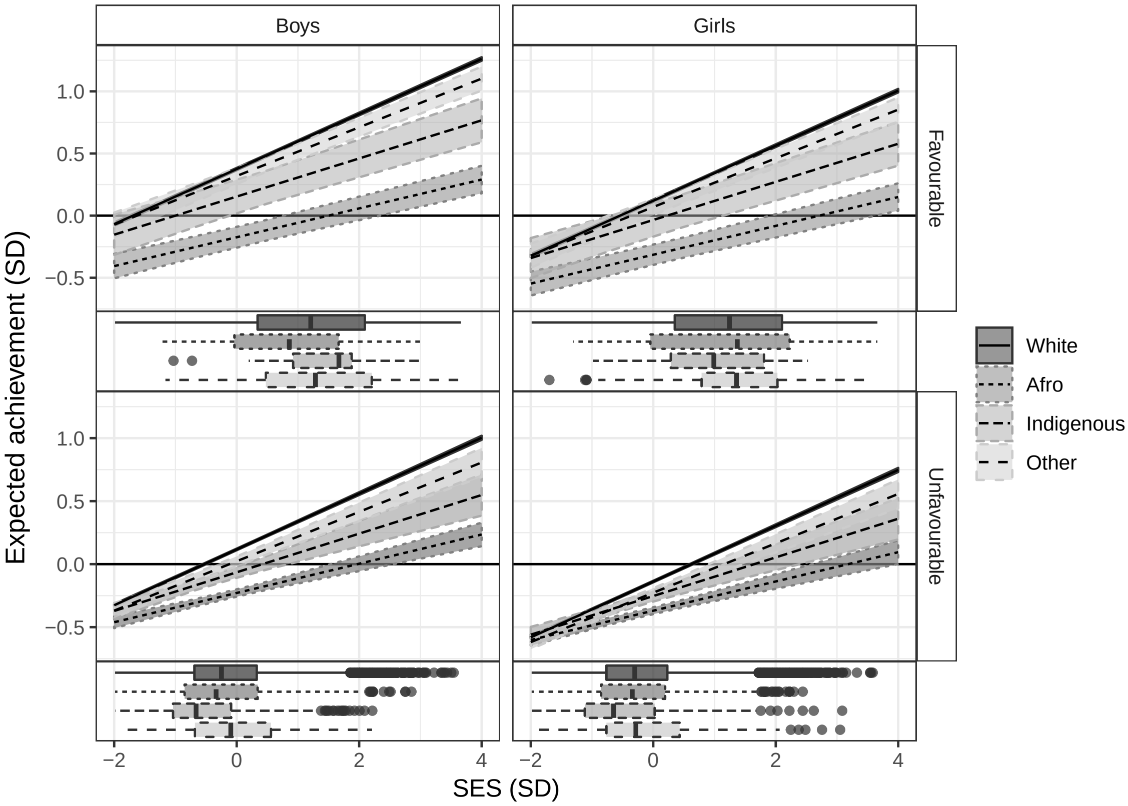

Figure 9.2: Predicted achievement for boys and girls of each ethnic group and school scenario by SES and SES distribution for each subgroup in 2008. The favourable scenario includes private schools with full-day schooling, and the unfavourable scenario includes state schools with morning schooling. In all scenarios it is assumed that students have average age and household size; schools are in an urban area, have an academic focus and a program that is not based on ethnoeducation; and LAs have average quality transfers and crime and conflict intensity.

of minority students are much more similar for those students in the 'unfavourable' than in the 'favourable' school scenario. The next section explores the differences in achievement between ethnic groups for subgroups of students in more detail.

9.4.2 Overall Achievement Gap

This section focuses on the estimated interaction terms for model (9.2), which allow studying how the overall ethnic achievement gap varies for different subgroups of students, schools and LAs. Figure 9.3 shows the estimated values and 95% confidence intervals for the interaction terms between the overall ethnic achievement gap and student and school characteristics[12] in the same model. In this figure, a positive estimated effect means that the overall ethnic achievement gap is wider for that particular subgroup of students, schools or LAs, while a negative effect means that the achievement gap is narrower for that subgroup. These comparisons are with respect to boys of average age and SES attending state schools offering full-day instruction.

As in chapters 6, 7 and 8, the results indicate the importance of studying the achievement gaps for each minority group. In this chapter, this means that student and school characteristics define different subgroups for students for each ethnic group. Figure 9.3 shows that SES defines subgroups of students with different achievement gaps for all minority groups. The conditional achievement gaps between White and Afrocolombian, Indigenous and other minority students are 0.11 SD 0.03 SD and 0.07 SD, respectively, wider for students with a SES one SD above the mean than for students with a mean SES.

Gender and school-day type only define subgroups of achievement gaps for Afrocolombian and Indigenous students, as shown in Figure 9.3. The achievement gaps between White and Afrocolombian and Indigenous students are 0.11 SD and 0.07 SD wider for boys than for girls. These gaps are also 0.18 SD and 0.11 SD narrower in schools offering morning school days and 0.2 SD and 0.13 SD narrower for Afrocolombian and Indigenous students, respectively, in schools offering afternoon school days[13], in comparison to schools offering full days.

Appendix A.7.3 shows that the students' age and attending a private school also define different subgroups of achievement gaps, although these results are not statistically significant in 2008. The achievement gap between White and Indigenous students is around 0.04 SD wider for students that are one SD above the mean age. Similarly, the achievement gap between White and Afrocolombian and Indigenous students is around 0.07 SD and 0.21 SD, respectively, wider for in state schools than in private schools.

Going back to the scenarios in section 9.4.1, Figure 9.4 presents the estimated overall ethnic achievement gap for boys and girls of different SES under the same 'favourable' and 'unfavourable' school scenarios, contrasting private schools with a full school day with state schools with a morning school day. The results show that those subgroups of student and schools that tend to obtain higher maths test scores are the same with

[12]Please note that no interaction was found with LA characteristics.

[13]The difference in the White-Indigenous gap between afternoon and full school day is not statistically different from zero in 2008 but it is for other years in the sample.

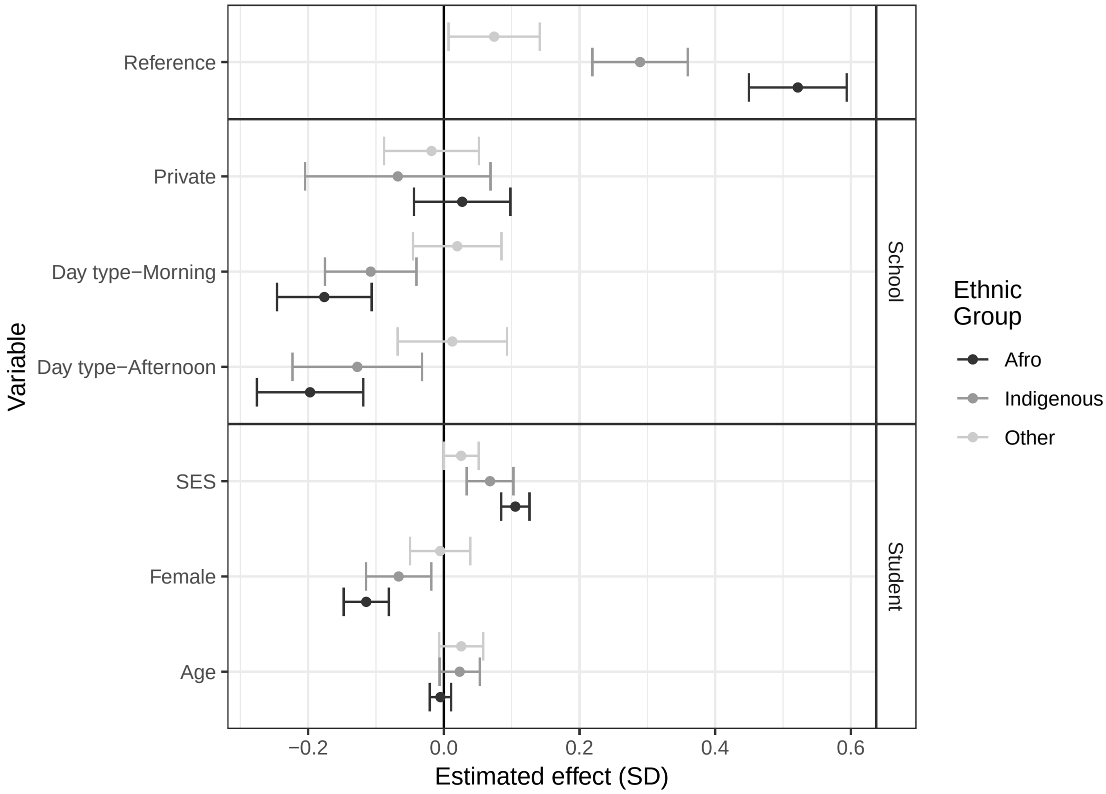

Figure 9.3: Estimated interaction effects and 95% confidence intervals for the single-level model including interactions (9.2) for maths achievement in 2008. The reference category includes boys of average age and SES attending state schools offering full-day instruction.

wider ethnic achievement gaps. Making the same comparison as in section 9.4.1, the achievement gaps between White and Afrocolombian, Indigenous and other minority girls with average SES in the 'unfavourable' school scenario are 0.23 SD, 0.11 SD and 0.09 SD, respectively. These gaps contrast with the estimated 0.55 SD, 0.22 SD and 0.06 SD achievement gaps between Afrocolombian, Indigenous and other minority and White students for boys with an average SES in the 'favourable' school scenario. Additionally, Figure 9.4 shows that, given the current distribution of SES, in none of these scenarios is it expected that minority students outperform White students, and the ethnic achievement gap is only statistically zero for students with a very low SES and therefore, low expected maths achievement.

Examining these interactions at each level of the education system (students, schools and LAs) sheds additional light upon the possible explanations for these patterns. This is explored in the next section.

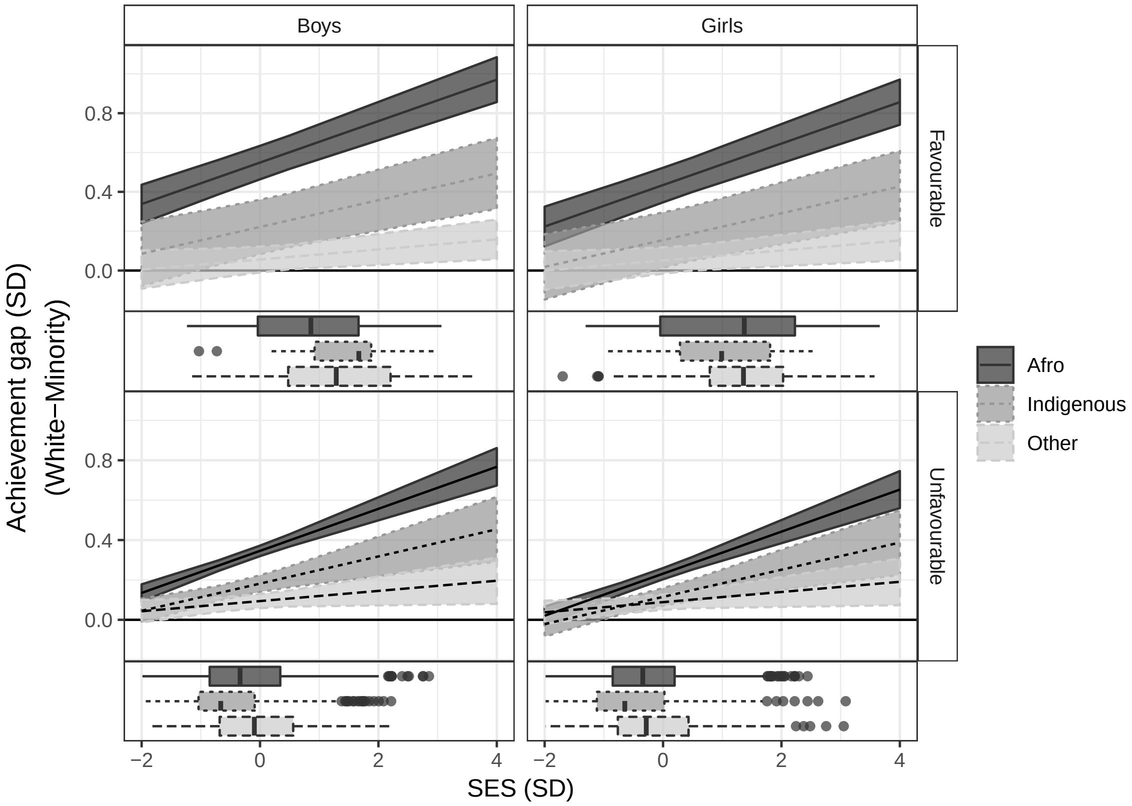

Figure 9.4: Estimated ethnic achievement gaps for boys and girls in each school scenario by SES and SES distribution for each subgroup in 2008. The favourable scenario includes private and have morning schooling, the unfavourable scenario includes state schools with morning schooling. In all scenarios it is assumed that students have average age.

9.4.3 Components of the Ethnic Achievement Gap

This section focuses on the results of estimating the three-level random intercept model (9.4), which allows exploring how the within-school ethnic achievement gap and the school and LA contextual effect of ethnicity vary among subgroups of students, schools and LAs. While the estimation results for the whole model are shown in appendix A.7.3, the following sections focus on the interaction terms for the within-school gap and the school and LA contextual effects. For reference the conditional within-school achievement gaps in 2008 are 0.1 SD, 0.05 SD and 0.05 SD for Afrocolombian, Indigenous and other minority students, respectively. The (negative) school contextual effects are 0.17 SD, 0.13 SD and 0.04 SD for the same groups. The LA contextual effects are 0.12 SD, 0.18 SD and 0.21 SD for Afrocolombian, Indigenous and other minority students, although none of these (detrimental) effects is statistically different from zero.

9.4.3.1 Within-school Achievement Gap

The within-school achievement gaps only interact with student characteristics: age, gender and SES. In other words, there is no evidence of a different within-school ethnic achievement gap for subgroups of schools or LAs, given the characteristics that are observed in the data[14]. Figure 9.5 shows the estimated interaction terms for the within-school achievement gap and their 95% confidence intervals. The reference category, in this case, represents the ethnic achievement gap for boys with an average age and SES.

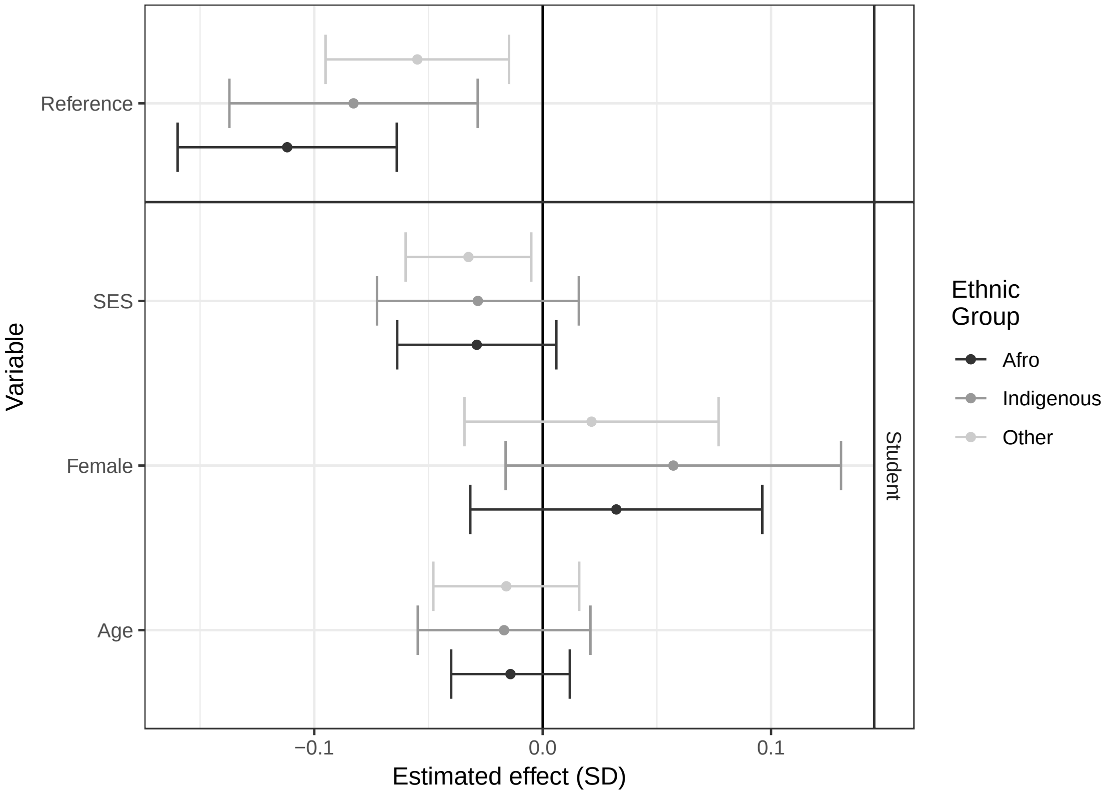

Figure 9.5: Estimated interaction effects and 95% confidence intervals for the within-school ethnic achievement gap according to the three-level random intercept model including interactions (9.4) for maths achievement in 2008. The reference category includes boys with an average age and SES.

In Figure 9.5, the only interaction term that is statistically significant is the within-school achievement gap for the group of other minority students and SES. This means that an increase of one SD in the SES is associated with a widening of 0.03 SD in the

[14]The results of the χ^2 test for the hypothesis that the within-school gap interacts with school and LA characteristics are presented in appendix A.7.1.

within-school ethnic achievement gap for students in the group of other minorities. This result also holds in 2010 and 2013, but in 2009, 2011 and 2012 the result, although similar, is not statistically significant.

Similarly, for other ethnic minority groups, the results, although consistent, are not always statistically significant, which may be associated with a lack of power given the small sample sizes for each specific subgroup. For example, Figure 9.5 shows that the within-school achievement gap for Afrocolombian students, is consistently estimated to be around 0.05 SD narrower for girls, but it is only statistically different from zero in 2013. For Indigenous students, an increase of one SD in the student's age is associated with a 0.03 SD widening in the within-school achievement gap, but this result is only statistically significant in 2011 and 2012.

To better visualise how these interactions imply different within-school ethnic achievement gaps, Figure 9.6 presents the estimated gap for boys and girls with different SES, as well as the SES distribution for these groups. It is assumed that all students have an average age. As shown in the figure, a higher SES is associated with a wider within-school ethnic achievement gap for all ethnic groups, but especially for the group of other minority students. The within-school ethnic achievement gap is also narrower for girls, but only the within-school achievement gap for Indigenous students is not statistically different from zero regardless of their SES.

Referring back to Figure 9.2, representing maths achievement for boys and girls of different SES and school scenarios, these results show that the within-school achievement gap is wider for those subgroups of students who tend to obtain higher test scores. Nonetheless, these gaps are the same regardless of the school and LA characteristics.

9.4.3.2 School Contextual Effect of Ethnicity

The school contextual effects of Afrocolombian, Indigenous and other minority students[15] vary among subgroups of students (by gender, age and SES) and schools (by school-day type). There are also interactions between the school and LA ethnic composition, which are described in section 9.4.3.3. Figure 9.7 figure presents the estimated interaction terms between the school contextual effects of ethnicity and these student and school characteristics and their 95% confidence intervals. The reference point is the

[15]As a reminder, the school contextual effect of ethnicity is the effect of an increase in the school proportion of minority students on the expected achievement, over and above the student's own ethnicity.

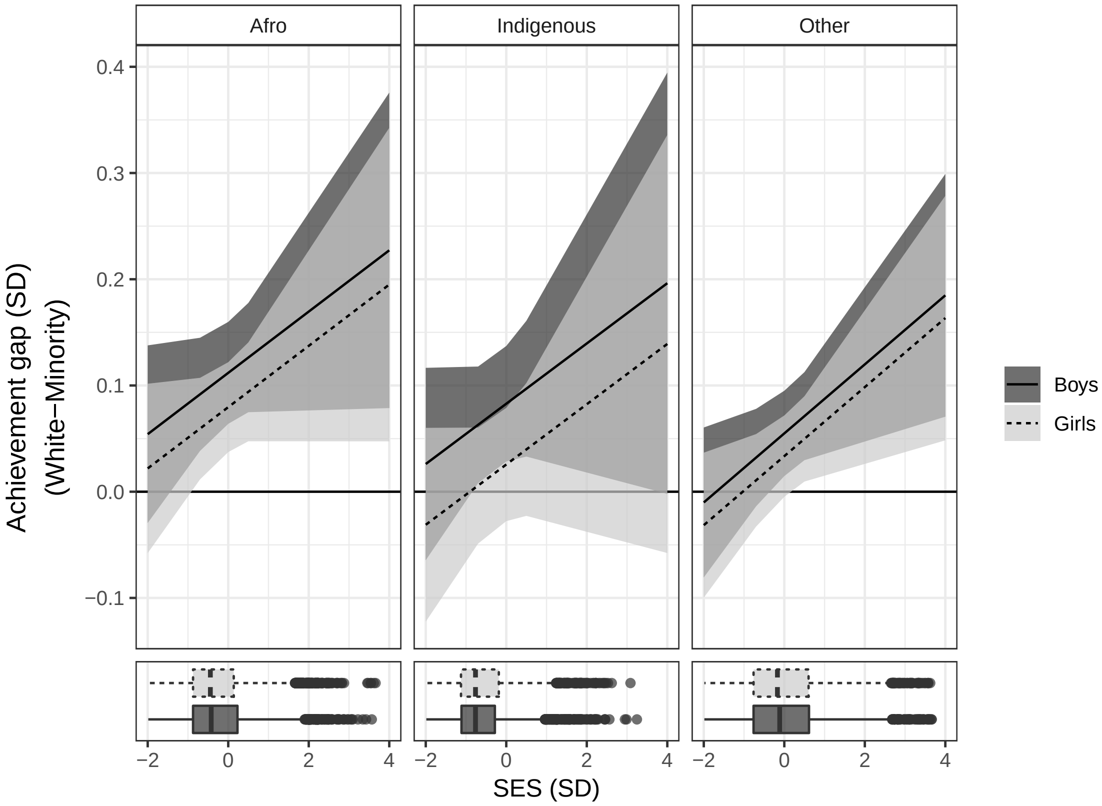

Figure 9.6: Estimated within-school ethnic achievement gaps for boys and girls by SES and SES distribution for each subgroup in 2008 It is assumed that students have average age.

school contextual effect for boys with average age and SES who attend schools offering full-day schooling in LAs with no minority students. Since the school contextual effect of Afrocolombian and Indigenous students is negative, a positive interaction term implies either a weaker (less negative) contextual effect of these ethnic groups or a switch from a negative to a positive contextual effect for that subgroup of students or schools, depending on the magnitude of the interaction term. In turn, a negative interaction term means a stronger (more negative) contextual effect for that subgroup.

As shown in Figure 9.7, while the negative contextual effects of Afrocolombian and Indigenous students vary across subgroups of students and schools, this is not the case for other minority students, whose school contextual effect is close to zero for all students and schools. In turn, the school contextual effects of Afrocolombian and Indigenous students are consistently 0.07 SD and 0.02 SD, weaker for girls than for boys, and 0.05 SD and 0.04 SD weaker for students one SD above the average age, than for stu-

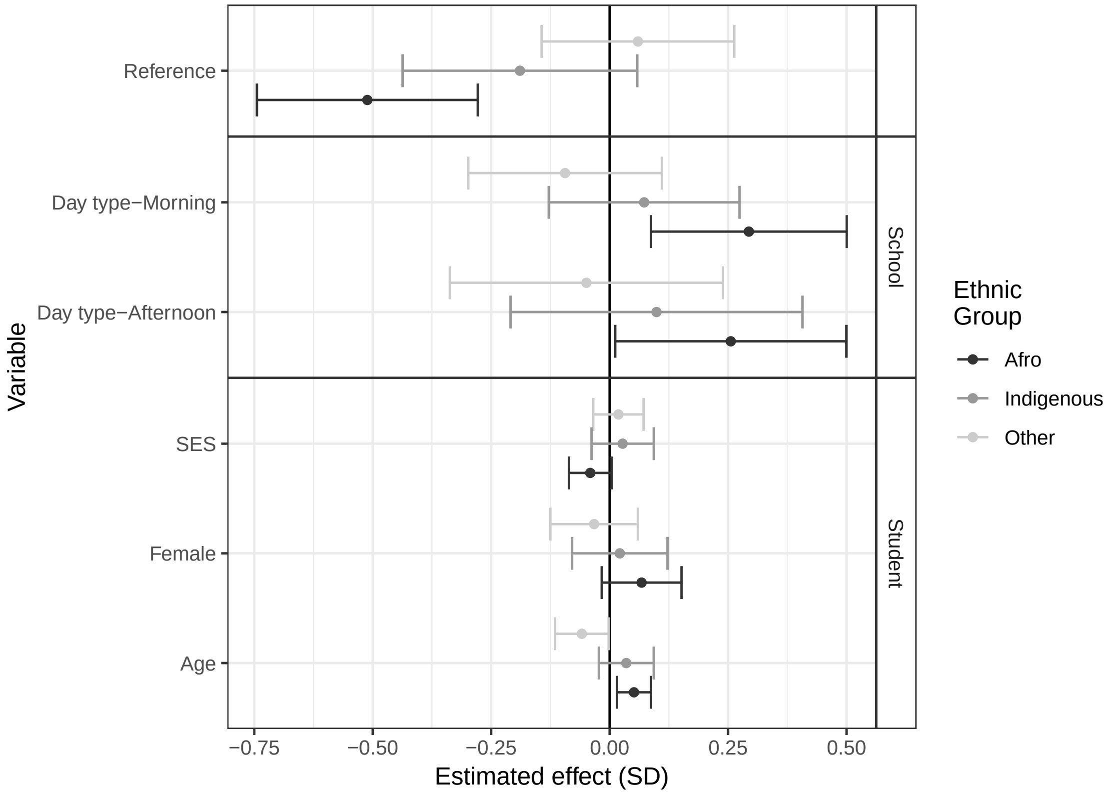

Figure 9.7: Estimated interaction effects and 95% confidence intervals for the school contextual effect of ethnicity according to the three-level random intercept model including interactions (9.4) for maths achievement in 2008. The reference point is the school contextual effect for boys with average age and SES who attend schools offering full-day schooling in LAs with no minority students.

dents with average age, respectively. Nonetheless, only the school contextual effect of Afrocolombian students is 0.05 SD stronger for each SD increase in the student's SES, although this interaction term is not statistically significant in 2008 and 2012, it is for other years in the sample.

The school contextual effects of Afrocolombian and Indigenous students also vary according to the school-day offered by the school, being stronger in schools offering full-day schooling. The school contextual effect of Afrocolombian students is 0.29 SD and 0.26 SD weaker in schools offering morning and afternoon school days, than full-day schooling, respectively. Similarly, the school contextual effect of Indigenous students is consistently 0.07 SD and 0.1 SD weaker for these schools. Although consistent over time, these differences are not statistically significant in 2008, but they are for all other years in the sample for schools offering morning instruction, and for 2009,2010 and 2012 for schools offering afternoon instruction.

To better understand how these estimated parameters imply different school contextual effects, Figure 9.8 presents the estimated school contextual effect for boys and girls of different SES in the favourable (full-day schooling) and unfavourable (morning schooling) school scenarios[16] presented in section 9.4.1. This figure shows that attending a school with a high proportion of Afrocolombian students is not always linked to poorer expected performance. In the unfavourable scenario, a higher proportion of Afrocolombian students is linked to higher average maths test scores, especially for students with a low SES. Similarly, the proportion of Afrocolombian students is only linked to poorer performance in the favourable school scenario for students a very high (2 SD above the mean) SES.

Once again, those subgroups of students that tend to have higher maths achievement, as shown in section 9.4.1, are also the subgroup of students for who the detrimental school contextual effect of ethnicity is stronger, especially the contextual effect of Afrocolombian students. Similarly, the contextual effect of ethnicity is more likely to be detrimental in the favourable than in the unfavourable school scenario.

9.4.3.3 Local Authority Contextual Effects of Ethnicity

The last component of the ethnic achievement gap, the LA contextual effect of ethnicity also varies according to student (age and gender) and school characteristics (school type – state or private- and the school's ethnic composition), but not by LA characteristics. Figure 9.9 shows the estimated interaction terms for the LA contextual effect of each ethnic group. The reference category corresponds to boys of average age attending private schools with no minority students.

The effect of living in a LA with a higher proportion of Afrocolombian and Indigenous students is around 0.1 SD and 0.19 weaker (less negative) for girls than for boys, respectively. However, in 2008 these figures are only 0.09 SD and 0.003 SD, respectively,

[16]Please note that these scenarios also assume that the LA proportion of minority students is zero for each ethnic group. This assumption facilitates the computation of the predicted scenarios and standard errors. As shown in section 9.4.3.3, this assumption is only relevant to the school contextual effect of Afrocolombian students, which is stronger (more negative) in LAs with a higher proportion of Indigenous students. In Figure 9.8 this would be visualised as a change in the intercepts of the lines representing the school contextual effect of Afrocolombian students.

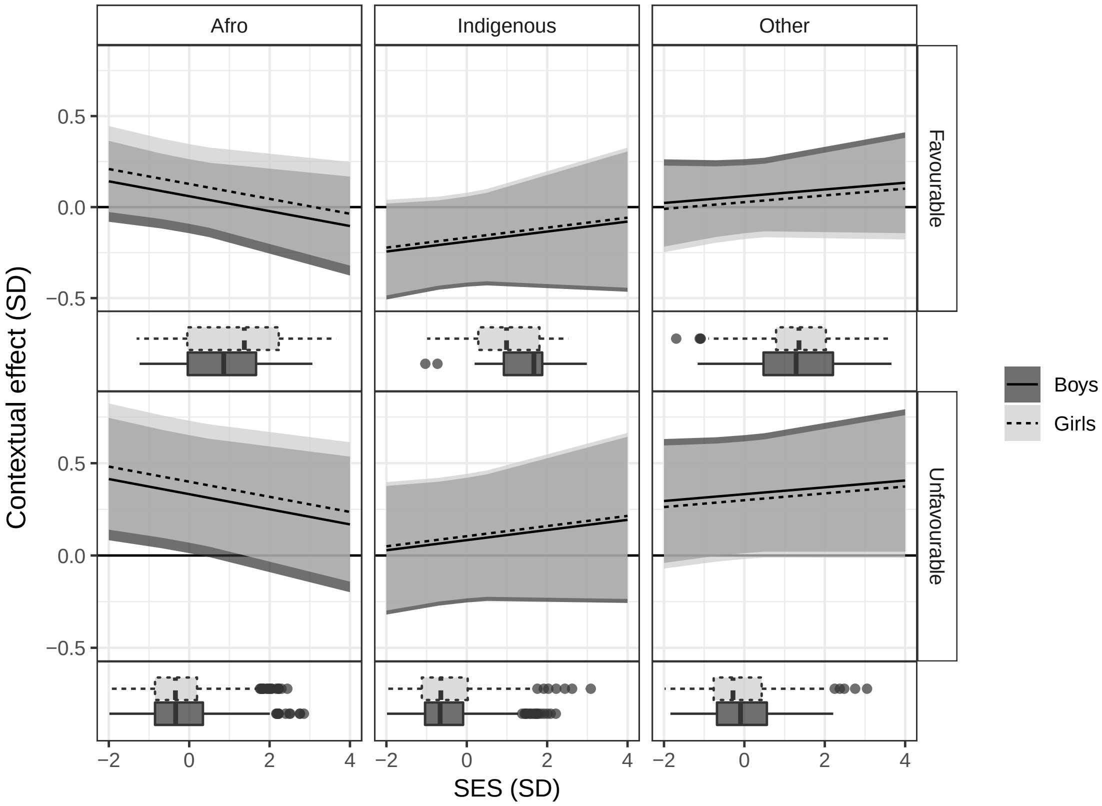

Figure 9.8: Estimated school contextual effect of ethnicity for boys and girls by SES and school scenario and SES distribution for each subgroup in 2008. The favourable scenario includes schools with full-day schooling, the unfavourable scenario includes schools with morning schooling. In all scenarios it is assumed that students have average age and LAs have no minority students.

as shown in Figure 9.9.

The only LA contextual effect that varies according to other student and school characteristics is that of Indigenous students. Such effect decreases (becomes more negative) 0.41 SD with each SD increase in the student's age. The negative effect of living in a LA with a higher proportion of Indigenous students is also 0.52 SD weaker (less negative) in private then in state schools and 0.22 SD stronger (more negative) in schools with a one percentage point higher proportion of Afrocolombian students. These two effects are not statistically significant in 2008, but they are for other years in the sample.

These interaction effects are better understood under the scenarios presented in Figure 9.10, which shows the LA contextual effect of each ethnic group according to the school proportion of Afrocolombian students[17], the students' gender and school sce-

[17]Please note that the boxplots in Figure 9.10 represent the distribution of students according to the pro-

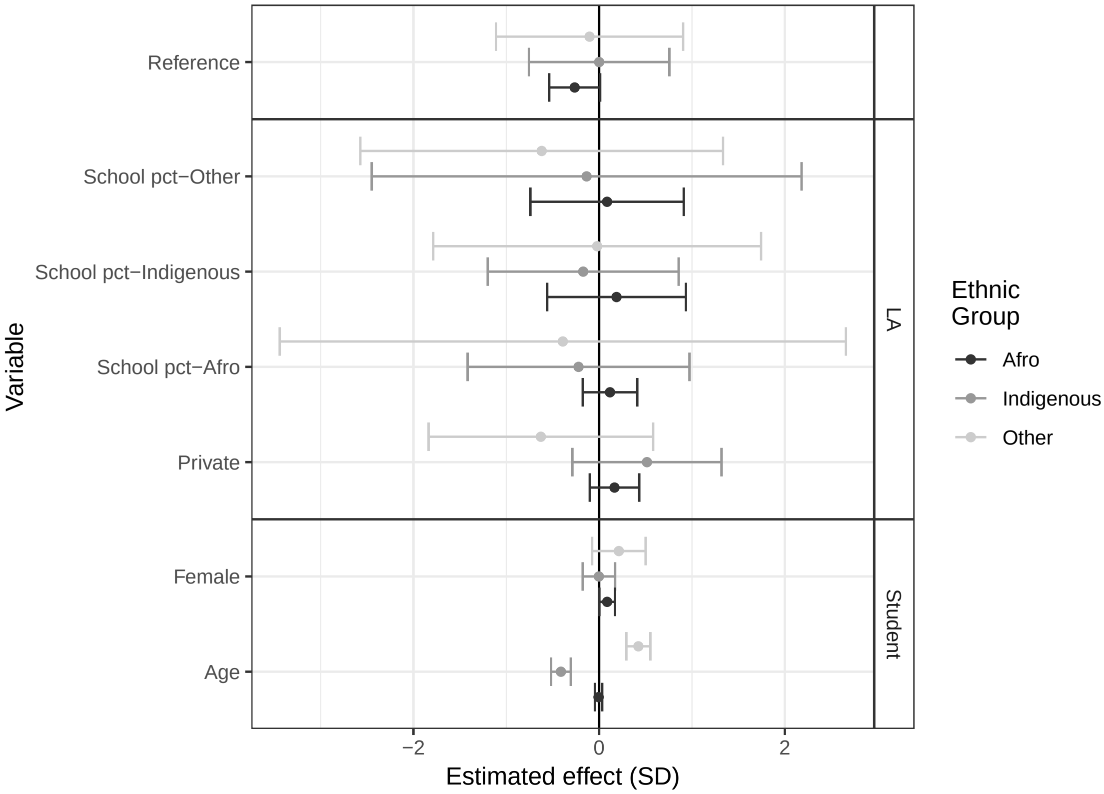

Figure 9.9: Estimated interaction effects and 95% confidence intervals for the LA contextual effect of ethnicity according to the three-level random intercept model including interactions (9.4) for maths achievement. The reference category corresponds to boys of average age attending private schools with no minority students.

nario (in this case, the favourable scenario indicates private schools and the unfavourable scenario indicates state schools). Although there is a high degree of uncertainty, the figure shows that under all scenarios, living in LAs with a higher proportion of minority students is linked to lower achievement, although this effect is slightly weaker for those attending private schools.

9.4.4 Random Variation in the Achievement Gap

Until now, the models have allowed the overall ethnic achievement gap and its components (the within-school gap and the school and LA contextual effect of ethnicity) to vary between subgroups of students, schools and LAs with known characteristics (e.g.

portion of Afrocolombian students in the schools they attend, and not the distribution of schools according to their proportion of Afrocolombian students. While there are enough student cases for the different scenarios, there are fewer schools in each of the scenarios.

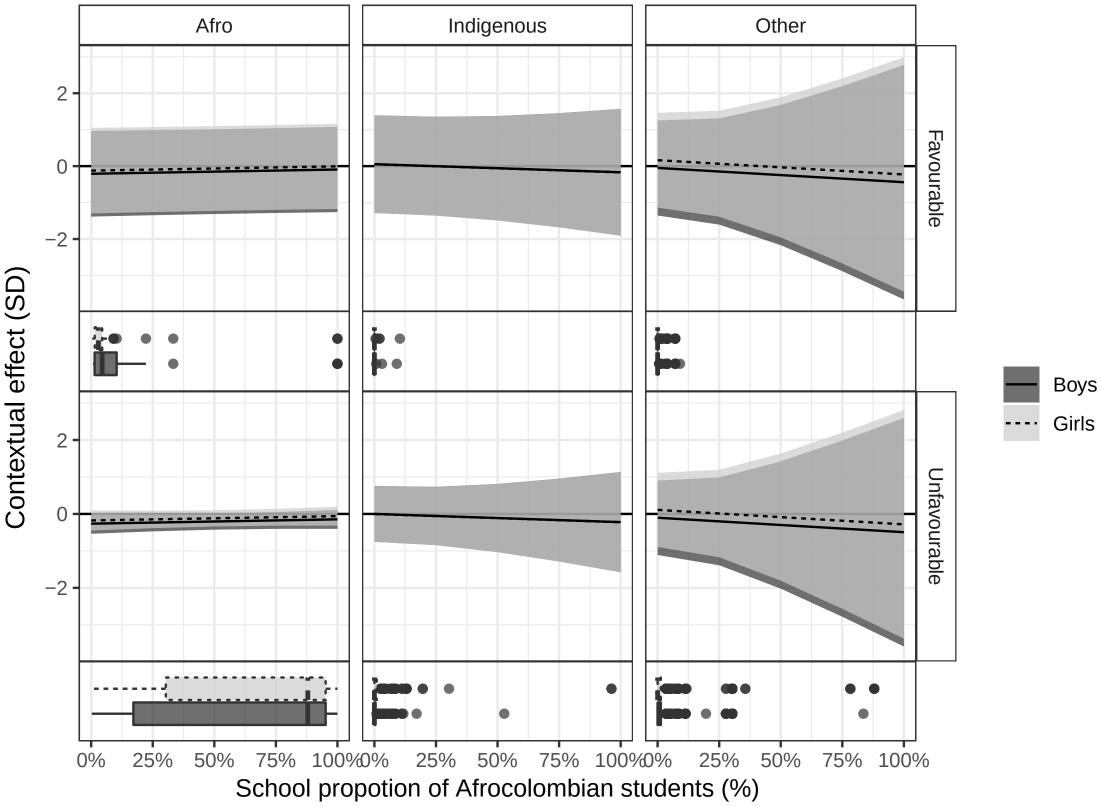

Figure 9.10: Estimated LA contextual effect of ethnicity for boys and girls by school ethnic composition and scenario and its distribution for each ethnic group in 2008. The favourable scenario includes private schools, the unfavourable scenario includes state schools. In all scenarios it is assumed that students have average.

gender, type, ethnic composition). This section studies models that allow the within-school achievement gaps to vary between schools[18] for reasons that are not observed in the data. The achievement gap may differ between schools either because of factors under their responsibility, such as their instructional and administrative practices or because of aspects outside their control, such as changes in the neighbourhood where the school is located or other types of random shocks.

In 2008, there is no evidence of random variation in the residual within-school ethnic achievement gap for Afrocolombian students between schools, as the school-level variance for the random slope $u_{A_{jk}}$ ($\sigma^2_{u_A}$) is estimated to be zero. This implies that the differences in average test scores between White and Afrocolombian students with sim-

[18]Models that allowed the within-school achievement gaps to vary between LAs randomly were also tested, but no evidence of random variation in the gap was found at this level, as shown in appendix A.7.1.

ilar characteristics and attending the same schools are the same across schools[19]

Nonetheless, the residual mean score for White students ($u_{1_{jk}}$) significantly differs from the national average in 1,008 schools and the residual within-school conditional ethnic achievement gaps for Indigenous ($u_{I_{jk}}$) and other minority students ($u_{O_{jk}}$) are significantly different to the national mean in 161 and 4 schools, respectively. Figure 9.11 shows the estimated random residual intercepts and within-school ethnic achievement gaps for each hundredth school in the sample in 2008, to improve the legibility of the graph.

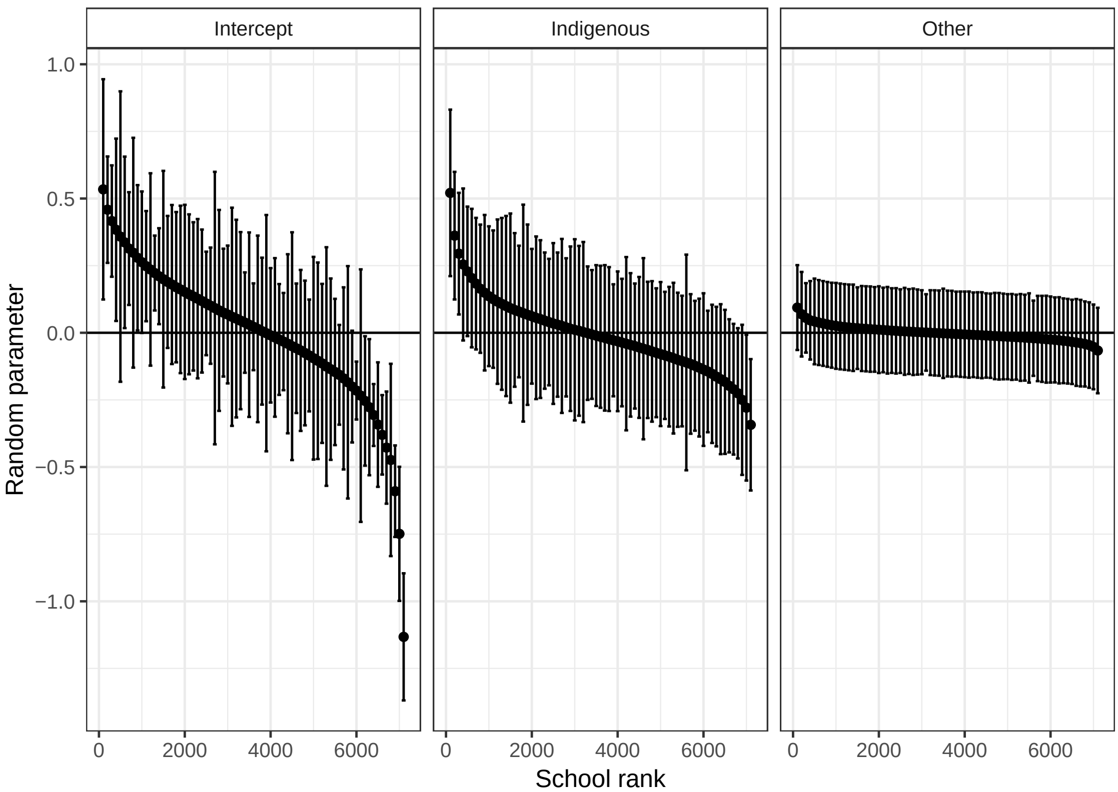

Figure 9.11: Caterpillar plot for the random residual intercept and within-school ethnic achievement gap of the model given by equations (9.4) and (9.5) in 2008. The plot shows each hundredth school in the sample to improve legibility.

Figure 9.12 displays the relationship between residual achievement for White students and the residual conditional within-school achievement gaps. A positive value of the within-school gap indicates that White students outperform minority students, while a negative value of this gap indicates that minority students outperform White

[19]Please note that there is no fixed between-school variation in the conditional within-school achievement gap either, as shown in section 9.4.3.1.

students in those schools. A positive value in the intercept indicates that White students in those schools perform above average, while a negative value indicates that these students perform below the average.

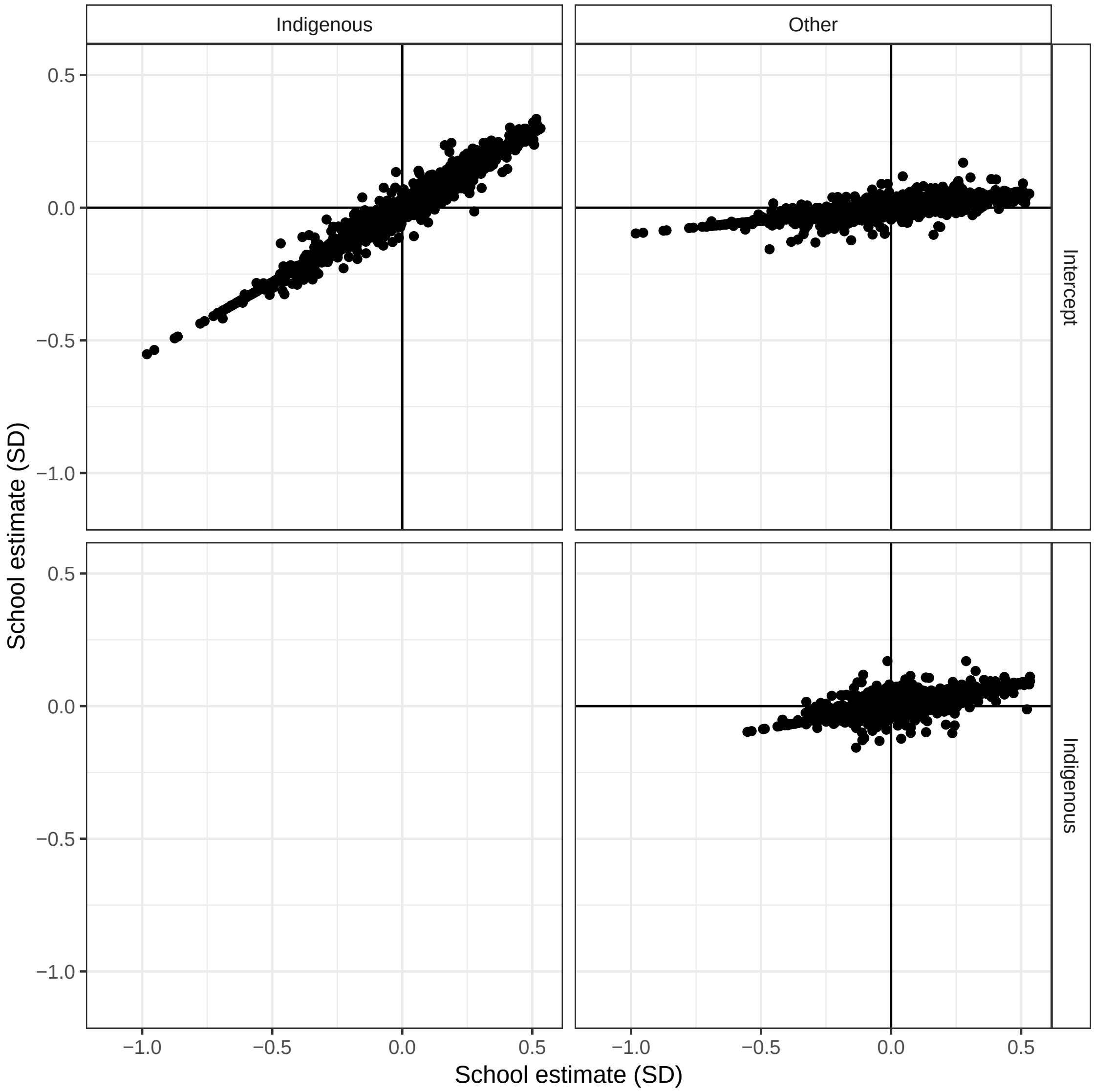

Figure 9.12: Scatter plots for the random residual intercept and within-school ethnic achievement gaps of the model given by equations (9.4) and (9.5) in 2008

According to Figure 9.12, there is a strong correlation between the residual achievement for White students and the residual within-school achievement gaps for Indigenous and other minority students. In 0.99% and 0.94% of the schools in which White students perform above the average, Indigenous and other minority students, respectively, have a worse performance than their White peers. In turn, in 0.99% and 0.96% of the schools in which White students perform below the average, Indigenous and other minority students, respectively, perform better than their White peers. Similarly, schools where Indigenous students outperform White students tend to be the same schools in which other minority students outperform White students.

Nonetheless, appendix A.7.3 shows that these results are not stable over time. For example, in 2009, the model estimates show no variation in the within-school gap for Indigenous students, and the random slope model fails to converge in 2010. As discussed in the next section, the random-part estimates are commonly found to be unstable over time (Bosker & Scheerens, 1989; Leckie & Goldstein, 2009; Perry, 2016).

9.5 Discussion

9.5.1 Summary of Findings

Consistent with Sánchez-Jabba (2011), this chapter finds that boys and students with higher SES attending urban schools and full school day tend to score higher on the SABER 11 maths test. These results hold for all ethnic groups. The general effects are also consistent with previous research on factors that affect academic achievement in Colombia (Abadía & Bernal, 2017; Gaviria & Barrientos Marín, 2001a; Soto Ceballos, 2014), as discussed in the context chapter (section 3.4.1).

Nevertheless, those variables that are related to a higher maths achievement are also related to wider overall ethnic achievement gaps, which are also consistent with evidence in the international literature that the ethnic gap is wider for boys (e.g. Bohrnstedt et al., 2015; Dekkers et al., 2000) and students with a higher SES (e.g. Biedinger et al., 2007; Dekkers et al., 2000; Reardon & Galindo, 2009; Strand, 2016). The school characteristics are much more country-specific, but the pattern is the same; schools in which students tend to get higher scores, such as private and full-day schools are also the schools with wider gaps. This pattern may arise due to floor effects (Zhu & Gonzalez, 2017). That is, that the students' scores do not fall below a threshold either because the questions are easy or, since the SABER 11 is a multiple-choice exam, due to the number of questions that students can correctly answer merely due to random chance. This would imply that there could be additional variation in the true achievement of students in these subgroups that the exams are unable to capture.

The finding that the overall ethnic achievement gaps interact with age, gender, SES, school day (morning, afternoon and full-day) and school type (state or private) is better understood when examining the components of the gap. Age and gender interact with all the components of the gap, which means that the overall gap is narrower for girls, for example, as the result of a combination of processes at the student, school and LA levels.

SES interacts with the within-school gap for the group of other minority students and the school contextual effect of Afrocolombian students, but not with the LA contextual effect of any ethnic group. In Colombia, Ramírez et al. (2012) argued that the rhetoric that justifies discrimination against ethnic minorities varies by SES, with biological explanations prevailing for minority people with a high SES, while cultural and historical arguments being applied to minority members with a low SES. This argument may help explain the wider achievement gaps for students of the group of other ethnic minorities with a higher SES. Similarly, Ramírez et al. (2012)'s reasoning may explain that the wider gap for Afrocolombian students is more likely to be linked to school-level than to student-level phenomena, which may include discriminatory practices that prevent schools with a high proportion of minority students from taking full advantage of a more favourable SES composition, but could also be linked to differences in the schools' policy and practice responses to a higher proportion of Afrocolombian students as their school intake becomes more socioeconomically advantaged. In turn, Voight et al. (2015)'s argument of differential school climate within schools seems to be a less likely explanation.

In contrast, school-day type only interacts with the school contextual effect of ethnicity, which means that the reason for schools with full-day schooling having wider achievement gaps than other schools might be linked to processes associated to the school ethnic composition, such changing pedagogical practices when the proportion of minority students increases (Sacerdote, 2011; Valoyes Chavez, 2015). Similarly, school type only interacts with the LA contextual effect of ethnicity, which implies that the reason why the overall ethnic achievement gap is narrower in private schools is not linked to a differential treatment of White and minority students within schools, as suggested in the international literature reviewed in chapter 2 (section 2.5.3) (e.g. Palardy et al., 2015; van Ewijk & Sleegers, 2010b). In turn, this finding suggests alternative explanations are more appropriate for the Colombian context. These might include the independence of private schools from the LA budgetary restrictions or other of the characteristics of those LA with a high proportion of minority students that are negatively linked to academic achievement.

The inclusion of random slopes for the within-school achievement gaps shows that the schools in which minority students tend to outperform White students are mainly those schools with below-average performance. This is a similar finding to the one reported by Muñoz-Chereau (2018) for the gender achievement gap in Chile. However, there is little residual variation in the within-school ethnic achievement gap after including all interaction terms, as Strand (2014a, 2014b, 2016) found in England.

Finally, unlike research in the UK and the US, no differences in the contextual effect of ethnicity were found by ethnic group (Canales & Webb, 2018; Hanushek et al., 2009; Hanushek & Rivkin, 2006; Noe et al., 2005; Ready & Reid, 2018). This implies that any policies aiming to reduce the negative effect of a high proportion of minority students would equally benefit students from all ethnic groups. This and other policy implications are further discussed in the next section.

9.5.2 Implications for Policy

The findings in this chapter reveal complex interlinks between academic achievement, the ethnic achievement gaps and student, school and LA characteristics. These complex interlinks have three main policy implications. First, that any policy to tackle ethnic achievement gaps should incorporate these complexities in their design, acknowledging that different subgroups of students, schools and LAs may require a differential treatment. For example, schools offering morning and afternoon school days may require a different treatment than schools with full-day schooling. At the same time, policy-makers should be vigilant for possible undesirable side effects of some interventions. For example, promoting full-day schooling may benefit some ethnic groups more than others, which may result in a higher level of inequality.

Second, the ethnic achievement gaps are wider for some subgroups of students and schools, such as for boys and state schools. This calls for prioritising interventions for these subgroups, after gaining a better understanding of what are the causal mechanisms behind the patterns observed in this chapter.

Finally, differences between students of the same ethnic group are as important as differences between White and minority students, especially for differences regarding SES, gender and school-day type (morning, afternoon or full-day) and zone (urban or rural). This presents a trade-off between inequality and achievement. From a policy perspective, both low inequality and high achievement are desirable outcomes. Whether the improvement in achievement is enough to compensate for the loss in equality is a matter of discussion for policymakers and society as a whole.

9.5.3 Limitations

Despite the robustness of most of this chapter's findings over time and for different sample specifications[20], as shown in appendix A.7, this study is limited in several ways.

First, appendix A.7.4 shows that the school-level residuals of the estimated random-slope models are not normally distributed, which is consistent over time and for the different subsamples that are tested for robustness. This is probably the result of a skewed distribution of maths test scores at the school level that is reflected by the distribution of the residuals. As Raudenbush and Bryk (2002) explained, the non-normality of the residuals does not bias the estimated coefficients but does their standard errors, which implies that inference tends to be over-optimistic. For that reason, the discussion focused on those findings that were found to be statistically significant in at least half of the observed years.

Second, a related concern is the statistical power to make comparisons between subgroups of students, schools and LAs. Given the number of interaction terms and control variables that are included in the models, these implicitly attempt comparisons that are only supported by a few observations in the data. Despite the large sample size that is available through the SABER 11 database, Colombia is highly segregated by SES and ethnicity, which makes some observations highly unlikely. For example, the number of White students who attend rural state schools with an ethnoeducation-based program and an afternoon school day varies between none in 2008 and 89 in 2012, while the model includes 68 coefficients to estimate (without considering their standard errors). The estimated parameters take the availability of information into account, but this compromises the ability to decide if an effect is due to random chance and, in some cases, leads to convergence problems in the estimation of multilevel models. The small number of observations for some instances also limits the complexity of the model that can be estimated. For example, the models in this chapter assume that the effect of SES

[20]These samples are i) consisting of schools that are observed for all years in the sample and ii) schools that serve students of at least two ethnic groups.

is the same regardless of gender or school type, and do not examine if the interaction terms are linked to student-level variables or their distribution among schools and LAs. Extending the models to consider further non-linearities requires a sample in which observations are more evenly distributed among these combinations of variables.

Third, caution should be exercised when using school estimates of random slopes to judge the performance of schools. Residuals capture any deviation from the predicted outcome, either under or outside the control of schools and are unstable over time. The lack of stability in the differential gap is consistent with concerns raised in the school effectiveness literature about the stability of value-added measures estimated from the residuals of multilevel models (e.g. Bosker & Scheerens, 1989; Gorard, 2010; Gorard, Hordosy, & Siddiqui, 2012; Leckie & Goldstein, 2009; Perry, 2016). For this reason, the discussion of the random variation in the ethnic achievement gap between schools is limited to establishing the degree of such variation and how it correlates with the residual achievement of White students.

Fourth, this chapter does not offer evidence of the causal mechanism behind the observed patterns. For example, it is not possible to differentiate if the variation in the achievement gap between subgroups of students occurs because of the characteristic of the subgroup (for example, because students receive full-day education) or because of confounding characteristics that are reflected by the grouping (for example, schools that offer full-day education are more successful in attracting and retaining better teachers or have access to more resources).

Finally, this chapter shares the scope, data and methods limitations with other research chapters, as discussed in sections 1.4, 4.6 and 5.4. Importantly, the data limitations imply that the subgroups that are studied in this chapter are limited to those that can be observed in the SABER 11 database. There is evidence, for example, of different ethnic achievement gaps in schools with different teacher's educational attainment (Stiefel et al., 2007), drop-out rates (Iv & Hogrebe, 2010) and student-sorting practices (Cebolla-Boado & Medina, 2011). None of these variables is observed in the data, and therefore, it is not possible to study if these findings also hold in the Colombian context. Nonetheless, similar groups may be characterised by private/state schools or schools with a different school day.

9.5.4 Implications for Future Research

An essential task for future research is to uncover the underlying causal mechanisms behind the patterns revealed in this chapter. Given the current policy environment in Colombia, an important task is to understand the reason why there is a wider achievement gap for schools with full-day schooling, as further discussed in section 10.4. This chapter showed that such explanations are probably linked to school-level processes rather than to student or LA processes.

The search for these explanations may be conducted using either qualitative research or taking advantage of differences in the periods and regions of the implementation of the expansion of full-day schooling, which may lead to natural experiments that allow identifying causal effects. More importantly, this search for causal effects should not only focus on how school-day type and other variables affect academic achievement, but on why these are different for each ethnic group.

This chapter also highlights the importance of approaching the analysis of ethnic achievement gaps as a multilevel phenomenon, understanding that the patterns are not only linked to the students' ethnicity but also to the schools' and LAs's ethnic composition. In the Colombian case, this approach allowed identifying that differences in the achievement gap by school type (state or private) are linked to LA-level processes, while differences by school-day type (morning, afternoon or full-day) are linked to school-level processes. Future research should consider incorporating this multilevel understanding of achievement gaps to the analysis of different contexts to advance towards a more holistic understanding of ethnic achievement gaps around the world.

9.6 Summary

This chapter built on the findings of chapter 8 and on Sánchez-Jabba (2011) to contribute to the ethnic achievement gap literature by:

1. Extending the time frame over which interactions with the ethnic achievement gap are analysed in Colombia, showing that student (age, gender and SES) and school characteristics (school type -state or private- and school-day type -morning, afternoon or full day-) clearly define subgroups of students with higher achievement and wider achievement gaps that are persistent over time.

2. Focusing on differences between groups instead of differential effects of variables related to achievement when interpreting interaction terms. This brings a new perspective to Sánchez-Jabba (2011)'s and others' interpretation as differential effects of these variables for minority students. This alternative focus allowed identifying subgroups of students, schools and LAs that should be a priority of future research and policy intervention to address the ethnic achievement gap.

3. Incorporating a multilevel perspective into the analysis by studying the within-school gaps and the school and LA effects of ethnicity, instead of focusing only on the overall achievement gaps. This shed additional light on the possible explanations for the existence of these interactions. For example, it showed that wider overall achievement gaps for schools offering full-day schooling are linked to school-level processes related to the school contextual effects of ethnicity, while narrower gaps for private schools are associated with LA processes, related to the LA contextual effects of ethnicity.

4. Exploring random between-school variation in the within-school achievement gaps after considering fixed variation according to student, school and LA characteristics, showing that minority students tend to outperform White students in those schools with a below-average performance for White students.

5. Illustrating a more holistic approach to the analysis of differences in the ethnic achievement gap by considering variation in the overall gap and its components according to observed student, school and LA characteristics but also unexplained variation, which although increasingly popular in the study of differential school effectiveness, has not been incorporated in the analysis of ethnic achievement gaps.

10 | Discussion and Conclusion

10.1 Introduction

Education, economics, sociology, human geography and other disciplines in the social sciences have studied ethnic achievement gaps, especially in the US and other industrialised countries. This thesis focused on two gaps in this literature: First, that the methods applied to study ethnic achievement gaps have often not been tailored with this aim in mind but have been borrowed from different areas of applications with little modification and second, that little is known about ethnic achievement gaps in developing countries such as Colombia.

This thesis tailored statistical techniques to measure and model ethnic achievement gaps in maths at the end of secondary education in Colombia between 2008 and 2013. Its findings contribute to the substantive and methodological literature on ethnic achievement gaps. They do so by approaching ethnic achievement gaps as a multilevel phenomenon, recognising the importance of studying each minority ethnic group separately (as opposed to pooled together as a whole) and elaborating on the methodological tools required for this approach. This thesis challenged the conventional view of the gaps as a student-level phenomenon that is explained by student, school or system-level processes. Instead, intricate patterns that arise not only from the student and school characteristics but also from their aggregation and links with different levels of the education system (students, schools and local authorities (LAs)) were studied.

This thesis includes four research chapters (chapter 6 to 9). Chapter 6 showed that overall ethnic achievement gaps in Colombia are not as wide as those reported in countries such as the US or the UK (Konstantopoulos et al., 2017; Leckie & Goldstein, 2019; Quinn & Le, 2018; Strand, 2016). Nonetheless, differences between schools and LAs with a different ethnic composition (school and LA contextual effects of ethnicity) are much more important in Colombia than in these other countries. This finding showed the importance of considering ethnic achievement gaps as a multilevel phenomenon; analysing all levels simultaneously as parts of a whole. Then chapters 7 and 8 showed that, however measured, socio-economic status (SES) is the most important explanation for ethnic achievement gaps in Colombia. Nonetheless, these chapters challenged the interpretation of previous studies, which view student-level processes as the main path through which SES (and other student characteristics) explains ethnic achievement gaps (e.g., Cook & Evans, 2000; Dustmann et al., 2010; Page et al., 2008; Sánchez-Jabba, 2011), by showing that, in Colombia, it is the link between school and LA SES composition and school and LA processes that explains most of the ethnic achievement gaps in Colombia. Finally, chapter 9 showed that SES is also linked to the type of school students attend. Only a small proportion of socioeconomically advantaged students are more likely to attend the best performing schools, which are also the schools with the wider ethnic achievement gaps. Students, although disproportionally White, belong to all ethnic groups.

This final chapter provides an overarching discussion of the findings presented in the research chapters (chapters 6, 7, 8 and 9). Each of these chapters presented an in-depth discussion for the specific literature to which they contributed. In turn, this chapter aims to link these findings to discuss how this thesis, as a whole, contributes to the ethnic achievement gap literature summarised in chapter 2 and its policy implications in Colombia. Section 10.2 presents the answers to this thesis' research questions, section 10.3 highlights the methodological and substantive contributions of answering these research questions and section 10.4 discusses their policy implications. Sections 10.5 and 10.6 point out this thesis' limitations and how future research could address them. Section 10.7 finalises with the concluding remarks.

10.2 Answers to the Research Questions

As section 1.2 introduced, this thesis aimed to answer four overarching research questions. The following summarises the answers to them.

1. **How to decompose the overall ethnic achievement gaps into differences within and between schools and LAs? What are the results of applying this technique to the Colombian context? - Chapter 6**

Mediation analysis can be used as a tool to decompose the ethnic achievement gaps into their components at different levels of the school system. It is enough to estimate a mediation model in which the school and LA ethnic composition mediate the relationship between the student's ethnicity and academic achievement. Translating the current methodological debate into the mediation analysis framework unlocks many potential extensions for the technique for decomposing the ethnic achievement gap, including adding levels of analysis and ethnic categories. Applying this technique to the Colombian context reveals small ethnic differences in achievement within schools, substantial differences in maths scores between schools and LAs with a different ethnic composition and high levels of segregation (measured as exposure). The combination of these factors implies that differences within schools are the smallest component of the ethnic achievement gaps, although the specific proportions vary among minority groups.

2. **How do the different index construction techniques for operationalising SES used in the literature affect the estimates of the conditional ethnic achievement gaps and their components at each level of the education system? - Chapter 7**

   The approaches to operationalise SES when using it as a control in a study of achievement gaps can be classified in five categories: using individual variables of education, occupation or income; using a variable that is employed for policy allocation (System for the Selection of Social Programs' Beneficiaries (Sisben) in Colombia); including all variables in the regression model or creating an index using prediction from a linear regression; creating an index using standardised variables (via factor analysis (FA), principal component analysis (PCA) or the sum/average of the variables); and creating an index without standardising the variables (such as the Hollingshead (2011)'s four-factor index). There are differences in the estimated ethnic inequalities in SES, ability to predict maths achievement and degree of within-school and LA clustering between these categories, but not much within them. These differences result in different estimated conditional ethnic achievement gaps and their components, especially for Indigenous students. It was argued that there different operationalisations may be preferable depending on the context and research questions.

3. **To what extent do student, school and LA characteristics explain the overall**

ethnic achievement gaps and their components at each level of the education system? How do single-level and multilevel modelling results compare when answering this question? - Chapter 8

Student, school and LA characteristics collectively explain up to 53%, 77% and 71% of the overall ethnic achievement gap between White and Afrocolombian, Indigenous and other minority students, respectively. Student-level characteristics explain the overall gap because of how they are distributed across schools and LAs; explaining the school and LA contextual effects of ethnicity. In contrast, they do not contribute to explaining the ethnic within-school gaps. Different school and LA characteristics contribute to explaining the school and LA contextual effects of ethnicity for each minority group. For Afrocolombian students, the most important explanations are LA characteristics such as their fiscal performance or crime and conflict intensity, while for Indigenous students it is the SES composition of LAs. When answering this question, the shrinkage of multilevel modelling is an advantage when studying the components of the gap, as they are less influenced by unreliable observations, but it is a disadvantage when studying the overall gap, as it is shrunk towards the within-school gap. In the last case, it is argued that single-level modelling is a preferable approach.

4. **Are Ethnic Achievement Gaps Uniform Across Subgroups of Students, Schools and Local Authorities? – Chapter 9**

   Ethnic achievement gaps tend to be narrower for those students, schools and LAs with characteristics that are associated with low achievement. Differences in the achievement gap by SES, gender and school-day type are particularly important, and this is mainly because the effect of attending schools with a higher proportion of minority students changes according to these characteristics. The within-school ethnic achievement gaps also tend to be narrower in those schools with a lower-than-expected average achievement, after considering all these characteristics and subgroups. These results seem to be the reflection of sorting effects associated with SES. Socioeconomically advantaged students of all ethnic groups tend to attend schools with above-average performance. Nonetheless, these are the schools for which the ethnic achievement gaps are wider.

10.3 Contributions

10.3.1 Methodological Contribution

This thesis contributed to tailoring existing statistical analysis methods for the study of ethnic achievement gaps as a multilevel phenomenon. In the literature, the need for a multilevel analysis is usually defended when analysing variables that are measured at different levels (e.g. students' achievement and school type). In this case, multilevel analysis is seen as a tool to avoid the ecological fallacy (interpreting relationships at the school-level as describing students) and the atomistic fallacy (assuming that student-level relationships also hold at the school-level) (e.g. Dansereau, Cho, & Yammarino, 2006; Hox, 2010; Johnston, Manley, Jones, Hoare, & Harris, 2017). The multilevel approach in this thesis shows that even when analysing the relationship between variables that are measured at the same level (e.g. students' achievement and ethnicity), these relationships hide intricate patterns involving different levels (e.g. students, schools and LAs). Analysing them shed additional light on this relationship, as it has been shown for the case of ethnic achievement gaps in Colombia.

Chapter 6 contributed to the methodological debate about the ethnic achievement gap decomposition into its components at different levels of the school system (Cook & Evans, 2000; Fryer & Levitt, 2004, 2006; Hanushek & Rivkin, 2006; Page et al., 2008; Reardon, 2008) by showing that mediation analysis can be used as a tool to extend the current methodology. In this thesis, the extensions considered additional levels of the education system (LAs in Colombia and the UK or school districts in the US) and multiple ethnic groups. It also illustrated the importance of these extensions by showing that the differences within schools are the least important component of the ethnic achievement gaps in Colombia, but their importance varies by ethnic group.

The methodological contribution made by chapter 7 is the comparison of widely used index-creation methods to operationalise SES for the study of ethnic achievement gaps. After identifying what these index-creation methods are, the chapter showed the importance of the different SES operationalisations for the estimation of conditional gaps is likely to be context-dependent. Thus, this chapter's contribution to the literature is the presentation of a framework to evaluate the effect of SES operationalisation, which includes assessing differences in the ethnic gaps in SES, predictive power for the outcome variable (e.g. maths achievement), within-school clustering and the estimated conditional ethnic achievement gaps and their components (the within-school gaps and the school and LA contextual effects of ethnicity). The application of this framework will enhance the transparency of future research on achievement gaps.

Chapter 8 contributed to the discussion of the advantages and disadvantages of using single-level and multilevel modelling when studying how observed variables can explain the ethnic achievement gaps. The discussion in this chapter showed that the use of multilevel modelling is not always preferable to the use of single-level models, as the choice depends on the specific research question and context. Again, this is a call for further reflection and transparency about how methodological decisions may influence the study's findings, and that the use of more sophisticated methods does not necessarily imply a better answer to the problem.

The overarching message of these methodological findings is that, when it comes to methodological approaches, there are no one-size-fits-all answers. Each method has advantages and disadvantages that need to be balanced against the specific research questions and context. The advice for researchers is then, first, to consider how competing methods can provide different insights for the problem at hand and, second, to be transparent about the consequences of their methodological choices for the conclusions of their study.

With this aim, this thesis has proposed four steps in the analysis of ethnic achievement gaps:

- As in chapter 6, identifying which level of the education system (e.g. students, schools and LAs) contributes the most and the least to the overall gaps. Here there is also room for choice between decomposition approaches, according to the relevant research question.

- As in chapter 7, evaluating and reporting how the operationalisation of key unobserved variables (SES in this case) can affect the conclusions of the analysis, assessing the criteria described above.

- As in chapter 8, examining how student, school and LA characteristics explain the overall ethnic achievement gaps and their components (the within-school gaps and school and LA contextual effects of ethnicity). This requires examining how the shrinkage of multilevel models affects the findings.

- As in chapter 9, considering interactions for the components of the gaps to examine how they and the overall gaps vary for subgroups of students, schools and LAs.

10.3.2 Substantive Contribution

This thesis contributed to a broader understanding of ethnic achievement gaps around the world and a more detailed understanding of ethnic achievement gaps in Colombia, adding to Sánchez-Jabba (2011), the only previous study about ethnic achievement gaps in Colombia. It did so by analysing the ethnic achievement gap as a multilevel instead of a student-level phenomenon, considering a different ethnic achievement gap for each minority group (instead of pooling all ethnic minorities as a single group) and exploring the consistency of the results over time instead of focusing on a single year.

Chapters 6, 8 and 9 showed that the distinctive features of each education system play a role in the existence of ethnic achievement gaps. Importantly, chapters 6 and 8 revealed that, in Colombia, student characteristics partially explain the ethnic achievement gap mainly throughout composition effects, and not because they represent a disadvantage for students per se, as current theories for the existence of these gaps suggest (Cervini et al., 2014; Conyers, 2002; Heuveline et al., 2010; Ogbu, 1995a; Phoenix & Husain, 2007). Similarly, chapter 9 showed that different ethnic achievement gaps for subgroups of student, schools and LAs result from interactions at different levels of the school system. For example, the gap is narrower for private schools because the LA contextual effect of ethnicity is narrower in these schools, and not because of student-specific responses. How these variables explain the gaps and define subgroups of students, schools and LAs was shown to vary among ethnic minorities, but to be consistent over time and robust to different sample specifications. These findings are possible thanks to the methodological contribution of this thesis, which reinforces the importance of re-thinking the ethnic achievement gap as a multilevel phenomenon, and not as one that can be uniquely attributed to either students or schools.

Throughout this thesis, there are recurrent reminders of the importance of segregation in explaining the ethnic achievement gaps in Colombia. Students of each ethnic group live in very different LAs, and the differences in the composition and characteristics of these LAs prove the most important explanations for the ethnic achievement gaps. Additionally, socioeconomically advantaged and disadvantaged students, who disproportionally belong to an ethnic minority -especially Indigenous-, are often concentrated in different schools, which also contributes to the ethnic achievement gaps. This thesis does not provide evidence that segregation is the cause of the problem; whether desegregation in itself would narrow the gaps is unknown. Even more, it may not be possible because the traditions and identity of some Afrocolombian and Indigenous communities are linked to their territories (Castillo Guzmán, 2008; Ferrero Botero, 2015).. Nevertheless, there is evidence that segregation is linked to differences in the access that White and minority students have to opportunities such as living in a LA that is less severely affected by conflict or attending a school with a full school day, and that the lack of access to these opportunities hinders the academic achievement of minority students.

10.4 Implications for Policy

As discussed in the context chapter (section 3.3.1.1), the most recent National Development Plan (NDP) in Colombia sets to target multiple inequalities, including ethnic inequality; recognising ethnic achievement gaps as a problem to be addressed (Congress of Colombia, 2019; DNP, 2018a). Although the evidence provided by this thesis is insufficient to comment on specific policies, it is possible to comment on three aspects of the NDP.

First, as discussed in section 10.3.2, chapters 8 and 9 of this thesis revealed that ethnic achievement gaps result from complex and overlapping processes linked to student, school and LA characteristics that are unique to each ethnic group. Therefore, as Rothstein (2004) argued for the US, any policy that aims to reduce the ethnic achievement gaps needs to incorporate interventions beyond the scope of educational policy. This thesis adds that it is also necessary to tackle inequalities among LAs with a different ethnic composition, as they proved especially important for the ethnic achievement gaps in Colombia. At the light of these findings, the fact that the NDP includes a differential policy line for each ethnic minority group for multiple aspects of ethnic inequality and has a regional focus (DNP, 2018a, p. 1157 - 1266) may be steps in the right direction to reducing ethnic achievement gaps. The key will be on the implementation of these plans, which, in order to reduce the ethnic achievement gaps, would need to prioritise the interventions that most effectively boost academic achievement in those LAs with a high proportion of minority students. Otherwise, the current plan risks perpetuating or widening the existing ethnic achievement gaps.

Second, chapter 8 showed that the achievement gaps are partially explained by ethnoeducation, mainly because schools with a higher proportion of minority students are more likely to offer a program based on ethnoeducation (educational programs for ethnic minorities) and these schools tend to obtain lower average SABER 11 maths scores. The ethnoeducation policy has been criticised for promoting ethnic segregation (Calvo Población & García Bravo, 2013). In fact, only 1.4% of schools in the country offer ethnoeducation programs and only 0.3%, 10.9%, 14.4% and 3.5%, of White, Afrocolombian, Indigenous and other minority 11th grade (age 16/17) students, respectively, attend schools offering ethnoeducation programs. Chapter 6 showed that the higher the segregation, the more important the differences between schools with varying ethnic composition when determining the ethnic achievement gap. As discussed in the context chapter (section 3.3.1.2), the NDP intends to strengthen the ethnoeducation policy. If strengthening the ethnoeducation program results in higher expected test scores in these schools, this policy may result in reduced ethnic achievement gaps. Please notice that offering an ethnoeducation program is also correlated with a socioeconomically disadvantaged student composition. Thus, it is unknown if strengthening ethnoeducation programs would result in improved academic achievement. Nonetheless, this thesis presents a call for questioning how to secure that the same opportunities are provided in schools offering ethnoeducation programs and other schools, if the policy continues to build on a segregated system (which should be evaluated).

Third, chapter 8 showed that ethnic achievement gaps are partially explained by the school-day type (morning, afternoon and full-day). Additionally, chapter 9 showed that the gaps are wider in schools offering full-day schooling than in schools offering morning or afternoon programs. The NDP's central educational policy is the expansion of full-day schooling as a path to increasing school quality. This thesis does not provide any evidence that full-day schooling *causes* wider achievement gaps or higher academic achievement, since schools offering this school day are different in many other aspects, including their student intake. Even more, since most of the effect of school-day type on the achievement gaps is transmitted through LA processes, it is possible that the findings in chapters 8 and 9 mainly reflect correlations between the availability of full-day schooling and other LA characteristics, such as the kind of educational policies that they adopt. Therefore, this thesis points out the need to better understanding the possible consequences of this policy regarding ethnic inequality before implementing it and in particular, paying attention to the LAs that are prioritised and how the full-day policy fits into their specific context.

As further discussed below, this thesis is limited in that it cannot offer specific interventions to narrow the ethnic achievement gap. Nevertheless, it offers support for holistic policies for ethnic minorities that consider the needs of each ethnic group. Any policy intervention also requires stepping back and thinking about how to tackle ethnic inequality, ideally without worsening the achievement of any student or damaging other outcomes of schooling besides maths test scores. Such a task requires the involvement of a broad set of actors, including students, teachers, parents, researchers, policymakers, leaders of ethnic minorities, educational leaders and the society in general.

10.5 Limitations

Beyond the specific limitations discussed in each research chapter (6, 7, 8 and 9), there are three overarching limitations of the evidence provided in this thesis. First, this thesis did not offer evidence of the causes of ethnic achievement gaps, despite uncovering complex links between student, school and LA characteristics behind the ethnic achievement gaps for each ethnic group. These links do not only reflect the effect of the characteristics that are observed in the data but also of all those other characteristics correlated with them. For example, although differences in SES explain a substantial proportion of the achievement gap for all minority groups, this not only reflects the effect of differences in education, occupation and income (used to measure SES) but also in parental practices or the type of role models available, for example. In this scenario, it is not clear whether the most effective policy recommendation is increasing the minimum wage to raise family income or promoting specific types of parental practice. Therefore, the lack of causal evidence limits the level of detail for policy recommendations that this thesis offers. What is clear, however, is that policies to tackle ethnic achievement gaps need to focus on reducing inequalities between LAs and schools with different ethnic composition.

Second, this thesis focused on maths academic achievement at the end of secondary education and did not consider other educational outcomes or stages of education. The focus improved the comparability of the findings in this thesis with the international literature. Like Sánchez-Jabba (2011), studies in the international literature (e.g., Page et al., 2008; Strand, 2016) often find that the overarching findings of ethnic achievement gaps hold for different subjects, although the magnitudes of the gaps may vary. Simultaneously, the lack of linked data for other educational stages hinders the analysis of the evolution of ethnic achievement gaps as students progress through the school system. Research has shown that ethnic achievement gaps appear in early childhood and increase over elementary schooling in the US (Davis-Kean & Jager, 2013; Quinn, 2015b) but gaps have narrowed or widened over different periods in the UK (Dustmann et al., 2010; Strand, 1999). Therefore, without further knowledge about the evolution of the ethnic achievement gaps over educational stages in Colombia, it is not possible to establish if the education system increases fairness in society or if it reproduces or propagates existing unequal structures.

Finally, this thesis' methodological approach did not consider the consequences of missing data, measurement error or more complex nesting structures. Although this thesis presented an innovative approach to analysing ethnic achievement gaps, the methods used can be further extended to better reflect the complexities of real data. Measurement error in ethnicity and the school and LA proportions of minority students would imply that the ethnic achievement gaps reported in this thesis are narrower than the actual achievement gaps (Goldstein, 1995). Measurement error in ethnicity may arise if not all minority students report belonging to an ethnic group or if not all students in the school or LA are observed in the SABER 11 dataset. (Goldstein, 1995) explained that the consequence of measurement error is that the estimated parameters are further away from zero than the real population parameter. Existing methods for tackling this problem require additional information that was not available at this moment (such as the reliability of the explanatory variables or alternative measures of ethnicity) (Carroll, Ruppert, Stefanski, & Crainiceanu, 2006). Similarly, missing data and a misspecified structure of the education system (ignoring classrooms and neighbourhoods, for example) may imply biases in the estimates of the ethnic achievement gap and its components. Nonetheless, robustness checks presented in the appendices A.4, A.6 and A.7 showed that the overarching findings in this thesis hold for different sample specifications. Therefore, the overarching findings of this thesis are likely to hold when accounting for the data characteristics described above.

Although current constraints did not allow solving these limitations, this thesis proposed a methodological approach that can be used to address them in future research, as discussed in the next section.

10.6 Directions for Future Research

This thesis proposed understanding the ethnic achievement gap as a multilevel phenomenon and outlines a methodological approach for analysing it, as further detailed in section 10.3.1. This can inform future substantive and methodological research. Substantively, researchers are encouraged to examine the contributions of different levels of education systems around the world in shaping ethnic achievement and other gaps, following the framework proposed in section 10.3.1. Additionally, researchers could explore further extensions to the methodological approach in this thesis, allowing for the study of data with more complex characteristics, including longitudinal, cross-classified and missing data and recognising measurement error.

Specifically for the Colombian context, this thesis pointed out the need for exploring the causal mechanisms behind the relationships between ethnicity and several student, school and LA characteristics. Some areas have been identified as a priority. First, understanding how ethnic segregation affects the within-school gap and school and LA contextual effects of ethnicity, and what are the potential benefits and drawbacks of promoting desegregation in Colombia. Second, identifying how the SES composition of LAs affects the ethnic achievement gap. There are many potential causal paths for this relationship, including the availability of jobs, attitudes toward ethnic minorities or differences in access to public utilities. Finally, evaluating the effects of current policies, such as the expansion of full-day schooling and strengthening ethnoeducation programs, on the ethnic achievement gaps.

These are complex tasks that require a joint effort from quantitative and qualitative researchers. Quantitative researchers can resort to causal inference methodologies, including searching for natural experiments and incorporating matching or instrumental variable approaches to get closer to identifying the unique contribution of observed characteristics. Qualitative researchers can then focus on understanding how these characteristics influence the gap. Identifying which of the possible alternatives exerts the strongest influence over academic achievement, especially for ethnic minorities can help inform better public policies to tackle the ethnic achievement gap.

10.7 Final Remarks

This thesis' overarching contribution to the literature was studying the methods that allow measuring and modelling the ethnic achievement gaps as a multilevel phenomenon and applying them to the Colombian context. Doing so has repercussions for practice and policy. In practice, researchers are encouraged to more carefully reflect on the methods they use to study the ethnic achievement gaps and to be more transparent about the effect of their methodological choices. For policy, this thesis expands the evidence base for the National Development Plan, which recognises the importance of reducing ethnic achievement gaps.

References

Abadía, L. K., & Bernal, G. (2017). A widening gap? A gender-based analysis of performance on the colombian high school exit examination. *Revista de Economia del Rosario*, *20*(1), 5–13. doi: 10.12804/revistas.urosario.edu.co/economia/a.6144

Agirdag, O., Demanet, J., Van Houtte, M., & Van Avermaet, P. (2011). Ethnic school composition and peer victimization: A focus on the interethnic school climate. *International Journal of Intercultural Relations*, *35*(4), 465–473. doi: 10.1016/j.ijintrel.2010.09.009

Agler, R., & De Boeck, P. (2017). On the interpretation and use of mediation: Multiple perspectives on mediation analysis. *Frontiers in Psychology*, *8*(NOV), 1–11. doi: 10.3389/fpsyg.2017.01984

Ainsworth-Darnell, J. W., & Downey, D. B. (1998). Assessing the Oppositional Culture Explanation for Racial/Ethnic Differences in School Performance. *American Sociological Review*, *63*(4), 536. Retrieved from `http://www.jstor.org/stable/2657266?origin=crossref` doi: 10.2307/2657266

Alviar, M., & Polanía, D. (1992). *Calidad de la educación secundaria en Colombia* (Tech. Rep.). Bogotá: Fedesarrollo.

Alwin, D. F., & Hauser, R. M. (1975). The Decomposition of Effects in Path Analysis. *American Sociological Review*, *40*(1), 37. Retrieved from `http://www.jstor.org/stable/2094445?origin=crossref` doi: 10.2307/2094445

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108. Retrieved from `http://dx.doi.org/10.1016/j.labeco.2014.05.008` doi: 10.1016/j.labeco.2014.05.008

Angulo Sálazar, R. C., García Cano, P., & Velasco Maldonado, J. (2009). *Manual operativo SISBEN III.* Bogotá: DNP. Retrieved from `https://www.sisben.gov.co/Portals/0/Documentos/DocumentosMetodologicos/01.02MANUALOPERATIVO_P.pdf`

APA Task Force on Socioeconomic Status. (2007). *Report of the APA Task Force on Socioeconomic Status* (Tech. Rep.). Washington, D.C: American Psychological Association. Retrieved from http://www.apa.org/pi/ses/resources/publications/task-force-2006.pdf doi: 10.1177/0898264310393339

Arellano, M. (1987). Computing Robust Standard Errors for Within-groups Estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–434. Retrieved from http://doi.wiley.com/10.1111/j.1468-0084.1987.mp49004006.x doi: 10.1111/j.1468-0084.1987.mp49004006.x

Arjona, A. (2016). *Rebelocracy*. Cambridge: Cambridge University Press. Retrieved from http://ebooks.cambridge.org/ref/id/CBO9781316421925 doi: 10.1017/9781316421925

Arteaga, I., & Glewwe, P. (2019). Do community factors matter? An analysis of the achievement gap between indigenous and non-indigenous children in Peru. *International Journal of Educational Development*, *65*(August), 80–91. Retrieved from http://dx.doi.org/10.1016/j.ijedudev.2017.08.003https://linkinghub.elsevier.com/retrieve/pii/S0738059317304959 doi: 10.1016/j.ijedudev.2017.08.003

Atal, J. P., Ñopo, H., & Winder, N. (2009). *New Century, Old Disparities.* Retrieved from http://www.iadb.org/document.cfm?id=2208929 doi: 10.1596/978-0-8213-8686-6

Atehortúa Cruz, A. (2006). La "revolución educativa": transcurso, resultados y perspectiva. *Análisis Político*, *19*(57), 126–152. Retrieved from http://www.revistas.unal.edu.co/index.php/anpol/article/view/46278

Atorresi, A., Pardo, C., Glejberman, D., Espinosa, G., Toranzos, L., Rocha, M., & Castro, M. (2010). *Compendio de los manuales del SERCE* (S. Block, Ed.). Santiago de Chile. Retrieved from http://unesdoc.unesco.org/images/0019/001919/191940s.pdf

Bali, V. A., & Alvarez, R. M. (2003). Schools and Educational Outcomes: What Causes the "Race Gap" in Student Test Scores? *Social Science Quarterly*, *84*(3), 485–507. doi: 10.1111/1540-6237.8403001

Bali, V. A., & Alvarez, R. M. (2004). The Race Gap in Student Achievement Scores : Longitudinal Evidence from a Racially Diverse School Disctrict. *The Policy Studies Journal*, *32*(3), 393–415.

Barbary, & Olivier. (2003). Sirven las categorias fenotipicas para entender la segregacion sociorracial en Cali, Colombia. In F. Lartigue & A. Quesnel (Eds.), *Las dinámicas de la población indígena: cuestiones y debates actuales en méxico* (pp. 149–187). Mexico: CIESAS. Retrieved from http://www.documentation.ird.fr/hor/fdi:010034231

Barbary, O., Ramírez, F., & Urrea, F. (2002). Identidad y ciudadanía afrocolombiana en la Región Pacífica y Cali: elementos estadísticos y sociológicos para el debate de la "cuestión negra" en Colombia. *Estudos Afro-Asiáticos*, *24*(3), 75–121.

Barrera-Osorio, F., García-Moreno, V., Patrinos, H. A., & Porta, E. (2011). Using the Oaxaca-Blinder Decomposition Technique to Analyze Learning Outcomes Changes over Time An Application to Indonesia's Results in PISA Mathematics. *Regional and Sectoral Economic Studies*, *11*(3), 65–84. Retrieved from http://www.usc.es/economet/journals2/eers/eers1134.pdf

Barrera-Osorio, F., & Gaviria, A. (2003). *Efficiency of colombian schools.* Bogotá. Retrieved from http://hdl.handle.net/11445/1364

Barrera-Osorio, F., Maldonado, D., & Rodríguez, C. (2012). *Calidad de la Educación Básica y Media en Colombia: Diagnóstico y Propuestas* (No. 126). Bogotá.

Barrientos Marín, J. H. (2008). Calidad de la educación pública y logro académico en Medellín 2004-2006. Una aproximación por regresión intercuartil. *Lecturas de Economía*, *68*(enero-junio), 121–144. Retrieved from http://www.banrep.gov.co/en/node/12993

Barrientos Marín, J. H., & Rios, P. (2007). Evaluación de la gestión privada del servicio público educativo en Medellín. *Lecturas de economia*, *66*(enero-junio), 147–172. Retrieved from http://www.scielo.org.co/pdf/le/n66/n66a5.pdf

Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of Multivariate Social Science Data* (Second ed.; A. Gelman, S. Rabe-Hesketh, J. S. Long, & A. Skrondal, Eds.). Boca Raton: Chapman and Hall/CRC. Retrieved from https://www.crcpress.com/Analysis-of-Multivariate-Social-Science-Data-Second-Edition/Bartholomew-Steele-Galbraith-Moustaki/p/book/9781584889601

Bartlett, M. S. (1937). The Statistical Conception of Mental Factors. *British Journal of Psychology. General Section*, *28*(1), 97–104. Retrieved from http://doi.wiley.com/10.1111/j.2044-8295.1937.tb00863.x doi: 10.1111/j.2044-8295.1937.tb00863.x

Bauer, T. K., & Sinning, M. (2008). An extension of the Blinder–Oaxaca decomposition to nonlinear models. *AStA Advances in Statistical Analysis*, *92*(2), 197–206.

Retrieved from http://link.springer.com/10.1007/s10182-008-0056-3 doi: 10.1007/s10182-008-0056-3

Becker, B. E., & Luthar, S. S. (2002). Social-emotional factors affecting academic outcomes among disadvantaged students: Closing the achievement gap. *Educational Psychologist*, *27*(4), 197–214. doi: 10.1207/S15326985EP3704

Begg, M. D., & Parides, M. K. (2003). Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Statistics in Medicine*, *22*(16), 2591–2602. Retrieved from http://doi.wiley.com/10.1002/sim.1524 doi: 10.1002/sim.1524

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & Quantity*, *53*(2), 1051–1074. Retrieved from https://doi.org/10.1007/s11135-018-0802-x doi: 10.1007/s11135-018-0802-x

Bell, A., & Jones, K. (2015). Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, *3*(01), 133–153. Retrieved from http://www.journals.cambridge.org/abstract_S2049847014000077 doi: 10.1017/psrm.2014.7

Bell, A., Jones, K., & Fairbrother, M. (2018). Understanding and misunderstanding group mean centering: a commentary on Kelley et al.'s dangerous practice. *Quality and Quantity*, *52*(5), 2031–2036. doi: 10.1007/s11135-017-0593-5

Bentaouet Kattan, R., Barrera-Osorio, F., Walter, A., & Taboada, B. (2008). *Colombia The Quality of Education in Colombia: An Analysis and Options for a Policy Agenda* (Tech. Rep.). Washington, DC: World Bank. Retrieved from http://documents.worldbank.org/curated/en/194851468025226653/Colombia-The-quality-of-education-in-Colombia-an-analysis-and-options-for-a-policy-agenda

BERA. (2011). *Ethical Guidelines for Educational Research.* London: Author. Retrieved from https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2011

Bidwell, C. E., & Kasarda, J. D. (1975). School District Organization and Student Achievement. *American Sociological Review*, *40*(1), 55. Retrieved from http://www.jstor.org/stable/2094447?origin=crossref doi: 10.2307/2094447

Biedinger, N., Becker, B., & Rohling, I. (2007). Early Ethnic Educational Inequality: The Influence of Duration of Preschool Attendance and Social Composition. *European*

Sociological Review, *24*(2), 243–256. Retrieved from https://academic.oup.com/esr/article-lookup/doi/10.1093/esr/jcn001 doi: 10.1093/esr/jcn001

Blanco, E. (2017). Los Alumnos Indígenas en México: Siete Hipótesis sobre el Rezago en los Aprendizajes de Nivel Primario (Indigenous Students in Mexico: Seven Hypothesis about the Learning Gap at the Primary Level). *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *15.3*(2017), 81–112. Retrieved from https://revistas.uam.es/index.php/reice/article/view/7966 doi: 10.15366/reice2017.15.3.005

Blanco álvarez, H. (2008). La integración de la etnomatemática en la etnoeducación. In *Encuentro colombiano de matemática educativa* (p. 7).

Bodovski, K., & Farkas, G. (2007). Do Instructional Practices Contribute To Inequality in Achievement?: the Case of Mathematics Instruction in Kindergarten. *Journal of Early Childhood Research*, *5*(3), 301–322. doi: 10.1177/1476718X07080476

Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., & Chan, D. (2015). *School Composition and the Black–White Achievement Gap* (Tech. Rep.). Washington, D.C: National Center for Education Statistics. Retrieved from https://eric.ed.gov/?id=ED560723

Bonilla Mejia, L. (2011). Doble jornada escolar y calidad de la educación en Colombia. *Coyuntura económica: Investigación económica y social*, *XLI*(1), 63–103.

Bornstein, M. H., Hahn, C.-S., Suwalsky, J. T. D., & Haynes, O. M. (2003). Socioeconomic status, parenting, and child development: The Hollingshead four-factor index of social status and the socioeconomic index of occupations. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 29–82).

Borooah, V. K., & Knox, C. (2015). Segregation, inequality, and educational performance in Northern Ireland: Problems and solutions. *International Journal of Educational Development*, *40*, 196–206. Retrieved from http://www.sciencedirect.com/science/article/pii/S0738059314000947 doi: 10.1016/j.ijedudev.2014.09.002

Bosker, R. J., & Scheerens, J. (1989). Issues in the interpretation of the results of school effectiveness research. *International Journal of Educational Research*, *13*(7), 741–751. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0883035589900256 doi: 10.1016/0883-0355(89)90025-6

Bourdieu, P. (1986). The Forms of Capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Westport, CT: Greenwood.

Bourdieu, P., & Passeron, J.-C. (1990). *Reproduction in Education, Society and Culture* (M. Featherstone, Ed.). London: SAGE Publications.

Braveman, P. A., Cubbin, C., Marchi, K., Egerter, S., & Chavez, G. (2001). Measuring socioeconomic status/position in studies of racial/ethnic disparities: maternal and infant health. *Public Health Reports*, *116*, 449–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12042609http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1497365

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling: An International Journal*, *1*(2), 103–124. Retrieved from http://journals.sagepub.com/doi/10.1177/1471082X0100100202 doi: 10.1177/1471082X0100100202

Buchmann, C. (2002). Measuring family background in international studies of education: Conceptual issues and methodological challenges. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 150–197). Washington, D.C: National Research Council. Retrieved from https://www.nap.edu/read/10322/chapter/7#151 doi: 10.17226/10322

Burgess, S., & Briggs, A. (2010). School assignment, school choice and social mobility. *Economics of Education Review*, *29*(4), 639–649. Retrieved from http://dx.doi.org/10.1016/j.econedurev.2009.10.011 doi: 10.1016/j.econedurev.2009.10.011

Burgess, S., & Greaves, E. (2013). Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities. *Journal of Labor Economics*, *31*(3), 535–576. Retrieved from http://www.jstor.org/stable/10.1086/669340 doi: 10.1086/669340

Burstein, L. (1980). The Analysis of Multilevel Data in Educational Research and Evaluation. *Review of Research in Education*, *8*, 158–23335.

Cafri, G., & Fan, J. (2018). Between-within effects in survival models with cross-classified clustering: Application to the evaluation of the effectiveness of medical devices. *Statistical Methods in Medical Research*, *27*(1), 312–319. Retrieved from http://journals.sagepub.com/doi/10.1177/0962280216628561 doi: 10.1177/0962280216628561

Cahan, S., Barneron, M., & Kassim, S. (2014). Gender differences in school achievement: a within-class perspective. *International Studies in Sociology of Education*, *24*(1), 3–23. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/09620214.2014.895132 doi: 10.1080/09620214.2014.895132

Caicedo, M. A., & Castillo Guzmán, E. (2011). "Spare the race and spoil the child" Racism, school books and afro-Colombian pedagogic writing. *Pedagogía y saberes*(34), 61–73.

Caicedo Ortíz, J. A. (2011). La cátedra de estudios afrocolombianos como proceso diaspórico en la escuela. *Peda*(34), 9–21.

Calvo Población, G. F., & García Bravo, W. (2013). Critical review of ethnic education in Colombia. *Historia y Educación Universidad de Salamanca*, *32*, 343–360.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249. Retrieved from http://www.tandfonline.com/doi/abs/10.1198/jbes.2010.07136 doi: 10.1198/jbes.2010.07136

Canales, A., & Webb, A. (2018). Educational Achievement of Indigenous Students in Chile: School Composition and Peer Effects. *Comparative Education Review*, *62*(2), 231–273. Retrieved from http://www.journals.uchicago.edu/doi/10.1086/696957 doi: 10.1086/696957

Cane, B. S. (1967). Educational research in England and Wales. *International Review of Education*, *13*(2), 152–161. Retrieved from https://link.springer.com/article/10.1007%2FBF01435889?LI=truehttp://link.springer.com/10.1007/BF01435889 doi: 10.1007/BF01435889

Cao, Z., & Maloney, T. (2018). Decomposing ethnic differences in university academic achievement in New Zealand. *Higher Education*, *75*(4), 565–587. doi: 10.1007/s10734-017-0157-6

Card, D., & Rothstein, J. (2007). Racial segregation and the black-white test score gap. *Journal of Public Economics*, *91*(11-12), 2158–2184. doi: 10.1016/j.jpubeco.2007.03.006

Caro, D. H., & Cortés, D. (2012). Measuring family socioeconomic status: An illustration using data from PIRLS 2006. In M. von Davier & D. Hastedt (Eds.), *Ieri monograph series: Issues and methodologies in large-scale assessments* (Vol. 5, pp. 9–33). Washington, D.C: IERInstitute. Retrieved from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_05_Chapter_1.pdf

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*. Chapman and Hall/CRC. Retrieved from https://www.taylorfrancis.com/books/9781420010138 doi: 10.1201/9781420010138

Casassus, J., Cusato, S., Froemel, J. E., Palafox, J. C., Willms, J. D., Sommers, A. M., & Pardo, C. (2001). *Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados para Alumnos del Tercer y Cuarto Grado de la Educación Básica. Informe Técnico*. Santiago de Chile: Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. Retrieved from http://unesdoc.unesco.org/images/0014/001492/149268s.pdf

Casey, L. (2016). *The Casey Review: A review into opportunity and integration* (Tech. Rep. No. December). London: Department for Communities and Local Government. Retrieved from https://www.gov.uk/government/publications/the-casey-review-a-review-into-opportunity-and-integration doi: December2016

Castellano, K. E., Rabe-Hesketh, S., & Skrondal, A. (2014). Composition, Context, and Endogeneity in School and Teacher Comparisons. *Journal of Educational and Behavioral Statistics*, *39*(5), 333–367. Retrieved from http://jeb.sagepub.com/cgi/doi/10.3102/1076998614547576 doi: 10.3102/1076998614547576

Castillo Guzmán, E. (2008). Etnoeducación y políticas educativas en Colombia: la fragmentación de los derechos. *Revista Educación y Pedagogía*, *XX*(52), 15–26. Retrieved from https://aprendeenlinea.udea.edu.co/revistas/index.php/revistaeyp/article/view/9879/9076

Causa, O., & Chapuis, C. (2011). Equity in Student Achievement Across OECD Countries. *OECD Journal: Economic Studies*, *2010*(1), 1–50. Retrieved from http://www.oecd-ilibrary.org/economics/equity-in-student-achievement-across-oecd-countries-an-investigation-of-the-role-of-policies_eco_studies-2010-5km61lb7b39x doi: 10.1787/eco_studies-2010-5km61lb7b39x

Cebolla-Boado, H., & Medina, L. G. (2011). The impact of immigrant concentration in Spanish schools: School, class, and composition effects. *European Sociological Review*, *27*(5), 606–623. doi: 10.1093/esr/jcq024

Cervini, R. A. (2006). Progreso de Aprendizaje en la Educación Secundaria Básica de Argentina: Un Análisis Multinivel de Valor Agregado. *REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *4*(3), 1–15.

Cervini, R. A. (2011). Equidad y oportunidad de aprender en la educación básica. *Profesorado*, *15*(nº 3, Diciembre), 20.

Cervini, R. A., Dari, N., & Quiroz, S. (2014). Estructura familiar y rendimiento académico en países de américa latina. *Revista mexicana de investigación educativa*, *19*, 569–597.

Chaparro, J. A. (2009). *"Es que tenía que ser negro": Estereotipos y relaciones sociales.* Texas. Retrieved from http://lanic.utexas.edu/project/etext/llilas/ilassa/2009/chaparro.pdf

Chapman, C., Muijs, D., Reynolds, D., Sammons, P., & Teddlie, C. (Eds.). (2016). *The Routledge International Handbook of Educational Effectiveness and Improvement Research, policy, and practice.* Abingdon: Routledge. Retrieved from https://www.routledge.com/The-Routledge-International-Handbook-of-Educational-Effectiveness-and-Improvement/Chapman-Muijs-Reynolds-Sammons-Teddlie/p/book/9780415534437

Charlton, C., Rasbash, J., Browne, W. J., Healy, M., & Cameron, B. (2019). *MLwiN Version 3.03.* Bristol: University of Bristol.

Chatterji, M. (2006). Reading Achievement Gaps, Correlates, and Moderators of Early Reading Achievement: Evidence From the Early Childhood Longitudinal Study (ECLS) Kindergarten to First Grade Sample. *Journal Of Educational Psychology*, *98*(3), 489–507. doi: 10.1037/0022-0663.98.3.489

Cheng, S., & Starks, B. (2002). Racial Differences in the Effects of Significant Others on Students' Educational Expectations. *Sociology of Education*, *75*(4), 306–327. Retrieved from http://www.jstor.org/stable/3090281

Chóez-Ortega, C. (2014). Causas de la discriminación racial en Latinoamérica: Reflexión teórica desde algunos conceptos sociales. *Revista Tecnológica ESPOL – RTE*, *27*(2), 105–121.

Christ, O., Hewstone, M., Schmid, K., Green, E. G., Sarrasin, O., Gollwitzer, M., & Wagner, U. (2017). Advanced multilevel modeling for a science of groups: A short primer on multilevel structural equation modeling. *Group Dynamics*, *21*(3), 121–134. doi: 10.1037/gdn0000065

Cofnas, N. (2016). Science Is Not Always "Self-Correcting". *Foundations of Science*, *21*(3), 477–492. Retrieved from http://link.springer.com/10.1007/s10699-015-9421-3 doi: 10.1007/s10699-015-9421-3

Coleman, J., Campbell, E. Q., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity* (Tech. Rep.). National Center for Educational Statistics. Retrieved from http://eric.ed.gov/?id=ED012275%5Cnhttp://files.eric.ed.gov/fulltext/ED012275.pdf doi: 10.2307/2091096

Collado, D., Lomos, C., & Nicaise, I. (2015). The effects of classroom socioeconomic composition on student's civic knowledge in Chile. *School Effectiveness and School Improvement*, *26*(3), 415–440. Retrieved from http://dx.doi.org/10.1080/09243453.2014.966725 doi: 10.1080/09243453.2014.966725

Colombia Aprende, & Ministry of Education. (n.d.). *Calendario Escolar en Colombia.* Retrieved 2016-03-31, from http://www.colombiaaprende.edu.co/html/home/1592/article-228192.html

Comfort, N. (2018). Genetic determinism rides again. *Nature*, *561*(7724), 461–463. Retrieved from http://www.nature.com/articles/d41586-018-06784-5 doi: 10.1038/d41586-018-06784-5

Condron, D. J. (2009). Social Class , School and Non-School Environments , and Black / White Inequalities in Children. *American Sociological Review*, *74*(5), 683.

Congress of Colombia. (1993). *Law 70 of 1993: By which the article 55 of the Constitution is developed (Afrocolombian studies).* Retrieved from http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=6943 doi: 10.1023/B:BUSI.0000032520.48012.fd

Congress of Colombia. (1994a). *Law 115 of 1994: By which it is issued the General Law of Education.* Retrieved from http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=292

Congress of Colombia. (1994b). *Law 142 of 1994, by which the public utilities regime is established.* Retrieved from http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=2752

Congress of Colombia. (2001). *Law 715 of 2001: By which organic rules regarding resources and competencies according to articles 151, 288, 356 and 357 of the Political Constitution and other regulations are dictated.* Retrieved from http://www.mineducacion.gov.co/1621/articles-86098_archivo_pdf.pdf

Congress of Colombia. (2002). *Law 732 of 2002, by which socioeconomic stratification is regulated.* Retrieved from http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=4736

Congress of Colombia. (2009). *Law 1324 of 2009: By which the parameter and criteria for organising the education quality assessment system are established, norms for the promotion of an assessment culture as a tool for State inspection and surveillance, and the Icfes is transformed.* Retrieved from http://www.secretariasenado.gov.co/

senado/basedoc/ley_1324_2009.html

Congress of Colombia. (2019). *Law 1955 of 2019: By which the National Development Plan 2018-2022: "Pact for Colombia, pact for equity" is created.* Retrieved from https://colaboracion.dnp.gov.co/CDT/Prensa/Ley1955-PlanNacionaldeDesarrollo-pacto-por-colombia-pacto-por-la-equidad.pdf

Connolly, P. (2006). The effects of social class and ethnicity on gender differences in GCSE attainment: a secondary analysis of the Youth Cohort Study of England and Wales 1997–2001. *British Educational Research Journal*, *32*(1), 3–21. doi: 10.1080/01411920500401963

Constitutional Court. (2013). *Auto 119 de 2013 de la Corte Constitucional* (Vol. 2004). Retrieved from http://www.corteconstitucional.gov.co/T-025-04/AUTOS2013/009Auto119de24dejuniode2013seguimientogobiernoreferentealcomponentederegistro.pdf

Conyers, J. (2002). Racial Inequality: Emphasis on Explainations. *The Western Journal of Black Studies*, *26*(4), 249–254.

Cook, M. D., & Evans, W. N. (2000). Families or Schools? Explaining the Convergence in White and Black Academic Performance. *Journal of Labor Economics*, *18*(4), 729–754. Retrieved from http://www.jstor.org/stable/10.1086/209975http://www.journals.uchicago.edu/doi/10.1086/209975 doi: 10.1086/209975

Correa, J. J. (2004). Determinantes del rendimiento educativo de los estudiantes de secundaria en Cali: un análisis multinivel. *Revista sociedad y economía*, *6*, 81–105.

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis : Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Education*, *10*, 1–9. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.9154&rep=rep1&type=pdf doi: 10.1.1.110.9154

Cowan, C. D., Hauser, R. M., Kominski, R. a., Levin, H. M., Lucas, S. R., Morgan, S. L., ... Chapman, C. (2012). *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A theoretical foundation* (Tech. Rep. No. November). Retrieved from http://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdfhttps://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf

Crawford, C. (2007). *When you are born matters : the impact of date of birth on Child Cognitive* *outcomes in England* (J. Payne, Ed.). London: The Institute of Fiscal Studies.

Cronbach, L. J., Deken, J. E., & Webb, N. (1976). *Research on Classrooms and Schools: Formulation of Questions, Design and Analysis.* (Tech. Rep.). Retrieved from https://eric.ed.gov/?id=ED135801

Cuevas Mendoza, M. L. (2013). *Sesgo Cultural en los ítems de las pruebas del examen SABER 11 en Colombia* (Masters Thesis). Universidad Nacional de Colombia.

DANE. (n.d.). *Estratifiación Socioeconómica para Servicios Públicos Domiciliarios.* Bogota. Retrieved from http://www.dane.gov.co/index.php/esp/estratificacion-socioeconomica/generalidades

DANE. (2005). *Análisis regional de los principales indicadores sociodemográficos de la comunidad afrocolombiana e indígena a partir de la información del censo general 2005* (Tech. Rep.). Bogotá: Author. Retrieved from http://www.dane.gov.co/files/censo2005/etnia/sys/Afro_indicadores_sociodemograficos_censo2005.pdf

DANE. (2007). *Colombia una nación multicultural* (Tech. Rep.). Retrieved from https://www.dane.gov.co/files/censo2005/etnia/sys/colombia_nacion.pdf

DANE. (2013). *Encuesta nacional de calidad de vida 2013: Presentación de resultados.* Bogota: Author. Retrieved from http://www.dane.gov.co/files/investigaciones/condiciones_vida/calidad_vida/Presentacion_ECV_2013.pdf

DANE. (2018). *Censo Nacional de Población y Vivienda 2018: Resultados preliminares.* Retrieved 2019-02-10, from https://sitios.dane.gov.co/cnpv-presentacion/src/#cuantos00

Dansereau, F., Cho, J., & Yammarino, F. J. (2006). Avoiding the "Fallacy of the Wrong Level". *Group & Organization Management*, *31*(5), 536–577. Retrieved from http://journals.sagepub.com/doi/10.1177/1059601106291131 doi: 10.1177/1059601106291131

Davis-Kean, P., & Jager, J. (2013). Trajectories of Achievement Within Race/Ethnicity: "Catching Up" in Achievement Across Time. *The Journal of Educational Research*, *107*(3), 197–208. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00220671.2013.807493 doi: 10.1080/00220671.2013.807493

De Melo, G. (2014). *Peer Effects Identified Through Social Networks: Evidence from Uruguayan Schools.*

Dee, T. S. (2005). A Teacher Like Me : Does Race, Ethnicity, or Gender Matter ? *The American Economic Review*, *95*(2), 158–165. Retrieved from http://www.jstor.org/

stable/4132809

de Jorge-Moreno, J., Díaz Castro, J., Rodríguez Vega, D. V., & Segura Gutiérrez, J. M. (2018, jan). Análisis de la eficiencia educativa y sus factores explicativos considerando el efecto de la titularidad en Colombia con datos Pisa 2012 (Analysis of educational efficiency and its explanatory factors considering the effect of ownership in Colombia with P. *Revista Desarrollo y Sociedad*, *80*(80), 89–118. Retrieved from https://revistas.uniandes.edu.co/doi/10.13043/DYS.80.3 doi: 10.13043/DYS.80.3

Dekkers, H. P. J. M., Bosker, R. J., & Driessen, G. W. J. M. (2000). Complex inequalities of educational opportunities: A large-scale longitudinal study on the relation between gender, social class, ethnicity and school success. *Educational Research and Evaluation*, *6*(March 2016), 59–82. doi: 10.1076/1380-3611(200003)6

del Cairo, C. L., & Jaramillo Marín, J. (2008). La etnicidad: Retos y perspectivas en la Colombia comtemporánea. *Novum*, *10*(28), 7–30.

Delgado Barrera, M. (2014). *La educación básica y media en Colombia: retos en equidad y calidad* (Tech. Rep.). Retrieved from http://www.fedesarrollo.org.co/wp-content/uploads/2011/08/La-educación-básica-y-media-en-Colombia-retos-en-equidad-y-calidad-KAS.pdf

Demack, S., Drew, D., & Grimsley, M. (2000). Minding the Gap: Ethnic, gender and social class differences in attainment at 16, 1988-95. *Race Ethnicity and Education*, *3*(2), 117–143. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/13613320050074005#.Vdq19rrufww doi: 10.1080/13613320050074005

Department for Education. (2014). *The national curriculum in England: Framework document* (Tech. Rep. No. December). London: Department for Education.

DNP. (2008a). *Conpes 117: Actualización de los criterios para la Determinación, Identificación y Selección de Beneficiarios de Programas Sociales* (Tech. Rep.). Bogotá: Departamento Nacional de Planeación. Retrieved from https://colaboracion.dnp.gov.co/CDT/Conpes/Social/117.pdf

DNP. (2008b). *Desempeño fiscal de los departamentos y municipios 2007* (Tech. Rep.). Bogotá: Author. Retrieved from https://colaboracion.dnp.gov.co/CDT/DesarrolloTerritorial/DESEMPE{~{N}}OFISCALULTIMO.pdf

DNP. (2013). *Base Certificada Sisben corte diciembre 2013.* Bogotá. Retrieved from https://www.sisben.gov.co/Información/ReporteBaseCertificadaSisbén

.aspx#.V5oOVvkrKUk

DNP. (2014). *índice de desempeño fiscal.* Bogotá: Author. Retrieved from https://www.dnp.gov.co/programas/desarrollo-territorial/Estudios-Territoriales/Indicadores-y-Mediciones/Paginas/desempeno-fiscal.aspx

DNP. (2018a). *Bases del Plan Nacional de Desarrollo 2018-2022: Pacto por Colombia, pacto por la equidad.* Bogotá: Author. Retrieved from https://www.dnp.gov.co/DNPN/Paginas/Plan-Nacional-de-Desarrollo.aspx

DNP. (2018b). *Transferencias del Sistema General de Participaciones.* Bogotá: Author. Retrieved from https://sicodis.dnp.gov.co/ReportesSGP/SGP_ReportePorEntidadyFuente.aspx

DNP, & Granados, M. d. P. (2003). *¿Quién se beneficia del SISBEN ? evaluación integral.* Bogotá: DNP.

Dollinger, S. J., & Clark, M. (2012). Test-Taking Strategy as a Mediator between Race and Academic Performance. *Learning and Individual Differences, 22*(4), 511–517. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1041608012000362 doi: 10.1016/j.lindif.2012.03.010

Duncan, B., & Duncan, O. D. (1968). Minorities and the Process of Stratification. *American Sociological Review, 33*(3), 356. Retrieved from http://www.jstor.org/stable/2091911?origin=crossref doi: 10.2307/2091911

Duncan, O. D., Cuzzort, R. P., & Duncan, B. (1961). *Statistical geography: Problems in analyzing areal data.* Glencoe: Free Press.

Dustmann, C., Machin, S., & Schönberg, U. (2010). Ethnicity and Educational Achievement in Compulsory Schooling. *The Economic Journal, 120*(546), F272–F297. Retrieved from https://academic.oup.com/ej/article/120/546/F272-F297/5089586 doi: 10.1111/j.1468-0297.2010.02377.x

Dworkin, A. G., & Turley, R. N. L. (2014). United States of America. In *The palgrave handbook of race and ethnic inequalities in education* (pp. 559–607). London: Palgrave Macmillan UK. Retrieved from http://link.springer.com/10.1057/9781137317803_19 doi: 10.1057/9781137317803_19

Education Endowment Foundation. (2019). *Teaching and Learning Toolkit.* Retrieved 2019-08-01, from https://educationendowmentfoundation.org.uk/

El Espectador. (2013, sep). *Icfes fijó fecha para pruebas Saber.* Bogotá. Retrieved from http://www.elespectador.com/noticias/nacional/icfes-fijo

-fecha-pruebas-saber-articulo-444854

Ellis, A. P. J., & Ryan, A. M. (2003, dec). Race and Cognitive-Ability Test Performance: The Mediating Effects of Test Preparation, Test-Taking Strategy Use and Self-Efficacy. *Journal of Applied Social Psychology*, *33*(12), 2607–2629. Retrieved from http://doi.wiley.com/10.1111/j.1559-1816.2003.tb02783.x doi: 10.1111/j.1559-1816.2003.tb02783.x

Elmelech, Y., & Lu, H.-H. (2004). Race, ethnicity, and the gender poverty gap. *Social Science Research*, *33*(1), 158–182. Retrieved from http://www.sciencedirect.com/science/article/pii/S0049089X03000449 doi: 10.1016/S0049-089X(03)00044-9

Elsevier B.V. (2017). *Scopus: Content Coverage Guide* (Tech. Rep. No. August). Retrieved from https://www.elsevier.com/__data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf

Ensminger, M. E., & Fothergill, K. (2003). A decade of measuring SES: What it tells us and where to go from here. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development.* (pp. 13–27). London. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2003-04572-001

ERIC. (1966). *Achievement.* Retrieved from https://eric.ed.gov/?qt=achievement&ti=Achievement

ERIC. (1980). *Educational Attainment.* Retrieved from https://eric.ed.gov/?qt=attainment&ti=Educational+Attainment

ESRC. (2015). *ESRC Framework for Research Ethics* (No. January). Retrieved from http://www.esrcsocietytoday.ac.uk/about-esrc/information/framework-for-research-ethics/index.aspx

Fairlie, R. W. (2005). An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement*, *30*(4), 305–316. Retrieved from https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JEM-2005-0259 doi: 10.3233/JEM-2005-0259

Fajardo M, D. (2015). *Estudio sobre los Orígenes del Conflicto Social Armado, Razones de su Persistencia y sus Efectos más Profundos en la Sociedad Colombiana* (Tech. Rep.). Bogotá: Espacio Crítico.

Fenton, S. (2010). *Ethnicity* (Second ed.). Cambridge: Polity Press. Retrieved from https://www.dawsonera.com:443/abstract/9780745675121

Ferrero Botero, E. (2015). Ethno-Education (Etnoeducación) in la Guajira, Colombia: Shaping Indigenous Subjectivities Within Modernity, Neoliberal Multiculturalism, and the Indigenous Struggle. *Latin American and Caribbean Ethnic Studies*, *10*(3), 288–314. Retrieved from http://dx.doi.org/10.1080/17442222.2015.1059542 doi: 10.1080/17442222.2015.1059542

Finch, F. H., & Hoehn, A. J. (1951). Measuring Socio-Economic or Cultural Status: A Comparison of Methods. *The Journal of Social Psychology*, *33*(1), 51–67. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00224545.1951.9921800 doi: 10.1080/00224545.1951.9921800

Fleischmann, F., Kristen, C., Heath, A. F., Brinbaum, Y., Deboosere, P., Granato, N., ... van de Werfhorst, H. G. (2014). Gender Inequalities in the Education of the Second Generation in Western Countries. *Sociology of Education*, *87*, 143–170. Retrieved from http://soe.sagepub.com/cgi/doi/10.1177/0038040714537836 doi: 10.1177/0038040714537836

Fordham, S., & Ogbu, J. U. (1986). Black students' school success: Coping with the "burden of 'acting white' ". *The Urban Review*, *18*(3), 176–206. Retrieved from http://link.springer.com/10.1007/BF01112192 doi: 10.1007/BF01112192

Formichella, M. M., & Ibañez Martín, M. (2014). Género e Inequidad Educativa: Un Análisis Para el Nivel Medio en Argentina. *Regional and Sectoral Economic Studies*, *14*(1), 195–2010. Retrieved from http://www.usc.es/economet/journals2/eers/eers14114.pdf

Francis, B., & Archer, L. (2005). British-Chinese pupils' and parents' constructions of the value of education. *British Educational Research Journal*, *31*(1), 89–108. doi: 10.1080/0141192052000310047

Fruehwirth, J. C. (2013). Identifying Peer Achievement Spillovers: Implications for Desegregation and the Achievement Gap. *Quantitative Economics*, *4*(1), 85–124. Retrieved from http://doi.wiley.com/10.3982/QE93 doi: 10.3982/QE93

Fryer, R. G., & Levitt, S. D. (2004). Understanding the Black-White Test Score Gap in the First Two Years of School. *The Review of Economics and Statistics*, *86*(May), 447–464. Retrieved from papers3://publication/uuid/477BD11C-0825-46DC-A529-BACADD200FC7 doi: 10.1162/003465304323031049

Fryer, R. G., & Levitt, S. D. (2006). The Black-White test score gap through third grade. *American Law and Economics Review*, *8*(2), 249–281. doi: 10.1093/aler/ahl003

Fryer, R. G., & Levitt, S. D. (2013). Testing for Racial Differences in the Mental Ability of Young Children. *American Economic Review*, *103*(2), 981–1005. Retrieved from http://pubs.aeaweb.org/doi/10.1257/aer.103.2.981 doi: 10.1257/aer.103.2.981

Fryer, R. G., & Torelli, P. (2010). An empirical analysis of 'acting white'. *Journal of Public Economics*, *94*(5-6), 380–396. Retrieved from http://dx.doi.org/10.1016/j.jpubeco.2009.10.011 doi: 10.1016/j.jpubeco.2009.10.011

Fuerst, J., Lynn, R., & Kirkegaard, E. (2019). The Effect of Biracial Status and Color on Crystallized Intelligence in the U.S.-Born African–European American Population. *Psych*, *1*(1), 44–54. Retrieved from https://www.mdpi.com/2624-8611/1/1/4 doi: 10.3390/Psychology1010004

Fulk, B. E., & Harrell, T. W. (1952). Negro-White Army test scores and last school grade. *Journal of Applied Psychology*, *36*(1), 34–35. Retrieved from http://content.apa.org/journals/apl/36/1/34 doi: 10.1037/h0061935

Gallo Restrepo, N., Meneses Copete, Y. A., & Minotta Valencia, C. (2014). Percepciones de las personas de ascendiencia africana en torno a la escuela, la iniversidad y la educación en Medellín-Colombia. *Cuadernos Interculturales*, *1*(22), 35–60.

Gamboa, L. F., Casas, A. F., & Piñeros, L. J. (2003). La teoría del valor agregado: una aproximación a la calidad de la educación en Colombia. *Revista de Economía del Rosario*, *6*(2), 95–116. Retrieved from https://revistas.urosario.edu.co/index.php/economia/article/view/1017/0

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard socio-economic index of occupational status. *Social Science Reserach*, *21*(1), 1–56.

Garbarski, D. (2015). Racial/ethnic disparities in midlife depressive symptoms: The role of cumulative disadvantage across the life course. *Advances in life course research*, *23*, 67–85. Retrieved from http://www.sciencedirect.com/science/article/pii/S1040260814000501 doi: 10.1016/j.alcr.2014.12.006

García Villegas, M., Espinosa Restrepo, J. R., Jiménez Angel, F., & Parra Heredia, J. D. (2013). *Separados y desiguales. Educación y clases sociales en Colombia*. Retrieved from http://www.dejusticia.org/files/r2_actividades_recursos/fi_name_recurso.591.pdf

Gaviria, A., & Barrientos Marín, J. H. (2001a). Calidad de la educación y rendimiento académico en Bogotá. *Coyuntura Social*(May), 111–127. Retrieved from http://

hdl.handle.net/11445/1759

Gaviria, A., & Barrientos Marín, J. H. (2001b). Características del plantel y calidad de la educación en Bogotá. *Coyuntura social*(25), 81–98.

Gaviria, A., & Barrientos Marín, J. H. (2001c). Determinantes de la calidad de la educación en Colombia. *Archivos de Economía*, *159*(1), 88. Retrieved from https://www.dnp.gov.co/portals/0/archivos/documentos/dee/archivos_economia/159.pdf

Gelman, A., & Park, D. K. (2009). Splitting a predictor at the upper quarter or third and the lower quarter or third. *American Statistician*, *63*(1), 1–8. doi: 10.1198/tast.2009.0001

Georges, A., & Pallas, A. M. (2010). New Look at a Persistent Problem: Inequality, Mathematics Achievement, and Teaching. *The Journal of Educational Research*, *103*(4), 274–290. doi: 10.1080/00220670903382996

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209–224. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0272775715300959 doi: 10.1016/j.econedurev.2016.03.002

Gillborn, D. (2006). Critical race theory and education: Racism and anti-racism in educational theory and praxis. *Discourse*, *27*(1), 11–32. doi: 10.1080/01596300500510229

Gillborn, D. (2008). Coincidence or conspiracy? Whiteness, policy and the persistence of the Black/White achievement gap. *Educational Review*, *60*(3), 229–248. doi: 10.1080/00131910802195745

Glewwe, P., Krutikova, S., & Rolleston, C. (2017). Do Schools Reinforce or Reduce Learning Gaps between Advantaged and Disadvantaged Students? Evidence from Vietnam and Peru. *Economic Development and Cultural Change*, *65*(4), 699–739. Retrieved from https://www.journals.uchicago.edu/doi/10.1086/691993 doi: 10.1086/691993

Goldstein, H. (1994). Multilevel Cross-Classified Models. *Sociological Methods & Research*, *22*(3), 364–375. Retrieved from http://journals.sagepub.com/doi/10.1177/0049124194022003005 doi: 10.1177/0049124194022003005

Goldstein, H. (1995). *Multilevel Statistical Models*. London. Retrieved from http://www.bristol.ac.uk/cmm/team/hg/msm-2nd-ed/

Goldstein, H., & Browne, W. J. (2005). Multilevel Factor Analysis Models for Continuous and Discrete Data. In *Contemporary psychometrics: a festschrift for roderick p. mcdonald* (pp. 453–475).

Gorard, S. (2010). Serious doubts about school effectiveness. *British Educational Research Journal*, *36*(5), 745–766. doi: 10.1080/01411920903144251

Gorard, S., Hordosy, R., & Siddiqui, N. (2012, nov). How Unstable are 'School Effects' Assessed by a Value-Added Technique? *International Education Studies*, *6*(1), 1–9. Retrieved from `http://www.ccsenet.org/journal/index.php/ies/article/view/22126` doi: 10.5539/ies.v6n1p1

Goussé, M., & Le Donné, N. (2014). Why Do Inequalities in 15-Year-Old Cognitive Skills Increased So Much in France? *Ssrn*, *1*(418). doi: 10.2139/ssrn.2533735

Greene, W. H. (2012). *Econometric Analysis* (7th ed.). New York, NY: Pearson.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The Achievement Gap and the Discipline Gap: Two Sides of the Same Coin? *Educational Researcher*, *39*(1), 59–68. Retrieved from `http://journals.sagepub.com/doi/10.3102/0013189X09357621` doi: 10.3102/0013189X09357621

Gregory, A., & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology*, *46*(4), 455–475. doi: 10.1016/j.jsp.2007.09.001

Grilli, L., & Rampichini, C. (2011). The role of sample cluster means in multilevel models a view on endogeneity and measurement error issues. *Methodology*. doi: 10.1027/1614-2241/a000030

Grogan-Kaylor, A., & Woolley, M. E. (2010). The social ecology of race and ethnicity school achievement gaps: Economic, neighborhood, school, and family factors. *Journal of Human Behavior in the Social Environment*, *20*(7), 875–896. Retrieved from `http://libproxy.wustl.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-23636-003&site=ehost-live&scope=site%5Cnagrogan@umich.edu` doi: 10.1080/10911359.2010.494927

Haag, N., Roppelt, A., & Heppt, B. (2015). Effects of mathematics items' language demands for language minority students: Do they differ between grades? *Learning and Individual Differences*, *42*, 70–76. Retrieved from `http://dx.doi.org/10.1016/j.lindif.2015.08.010` doi: 10.1016/j.lindif.2015.08.010

Hampden-Thompson, G., Guzman, L., & Lippman, L. (2013). A cross-national analysis of parental involvement and student literacy. *International Journal of Comparative*

Sociology, *54*(3), 246–266. Retrieved from http://journals.sagepub.com/doi/10.1177/0020715213501183 doi: 10.1177/0020715213501183

Hanushek, E. A. (1970). *The Production of Education, Teacher Quality and Efficiency* (Tech. Rep.). Washington, D.C: Bureau of Educational Personnel Development. Retrieved from https://eric.ed.gov/?id=ED037396

Hanushek, E. A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*, *19*(2), 141. Retrieved from http://links.jstor.org/sici?sici=0162-3737%28199722%2919%3A2%3C141%3AATEOSR%3E2.0.CO%3B2-3&origin=crossref doi: 10.2307/1164207

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2009). New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement. *Journal of Labor Economics*, *27*(3), 349–383. Retrieved from http://www.journals.uchicago.edu/doi/10.1086/600386 doi: 10.1086/600386

Hanushek, E. A., & Rivkin, S. G. (2006). *School quality and the black-white achievement gap* (No. October). Cambridge, MA. Retrieved from http://www.nber.org/papers/w12651

Harris, D. N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *The Teachers College Record*, *112*(4), 1163–1197. doi: 10.1017/CBO9781107415324.004

Hart Reyes, L., & Stanic, G. M. A. (1988). Race, Sex, Socioeconomic Status, and Mathematics. *Journal for Research in Mathematics Education*, *19*(1), 26–43. Retrieved from http://www.jstor.org/stable/749109?origin=crossref doi: 10.2307/749109

Hartas, D. (2015). Patterns of Parental Involvement in Selected OECD Countries: Cross-National Analyses of PISA. *European Journal of Educational Research*, *4*(4), 185–195. doi: 10.12973/eu-jer.4.4.185

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*(3), 451–470. Retrieved from http://doi.wiley.com/10.1111/bmsp.12028 doi: 10.1111/bmsp.12028

Hellebrandová, K. (2014). El proceso de etno-racialización y resistencia en la era multicultural: Ser negro en Bogotá. *Universitas Humanística*, *77*(77). Retrieved from http://revistas.javeriana.edu.co/index.php/univhumanistica/article/view/5934 doi: 10.11144/Javeriana.UH77.eper

Herman, M. R. (2009). The Black-White-Other Achievement Gap: Testing Theories of Academic Performance Among Multiracial and Monoracial Adolescents. *Sociology of Education*, *82*(1), 20–46. Retrieved from http://journals.sagepub.com/doi/10.1177/003804070908200102 doi: 10.1177/003804070908200102

Herrnstein, R. J. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Heuveline, P., Yang, H., & Timberlake, J. M. (2010). It takes a village (perhaps a nation): Families, states, and educational achievement. *Journal of Marriage and Family*, *72*(5), 1362–1376. doi: 10.1111/j.1741-3737.2010.00770.x

Heyneman, S. P. (1976). Influences on Academic Achievement: A Comparison of Results from Uganda and More Industrialized Societies. *Sociology of Education*, *49*(3), 200–211. Retrieved from http://www.jstor.org/stable/2112231?origin=crossref doi: 10.2307/2112231

Heyneman, S. P., & Loxley, W. A. (1983). The Effect of Primary-School Quality on Academic Achievement Across Twenty-nine High-and Low-Income Countries. *American Journal of Sociology*, *88*(6), 1162–1194.

Hollingshead, A. B. (2011). Four factor index of social status (Unpublished Working Paper, 1975). *Yale Journal of Sociology*, *8*, 21–52. Retrieved from http://elsinore.cis.yale.edu/sociology/yjs/yjs_fall_2011.pdf#page=21

Hou, F. (2014). A General Approach to Effect Decomposition. *Social Science Quarterly*, *95*(3), 894–904. Retrieved from http://doi.wiley.com/10.1111/ssqu.12074 doi: 10.1111/ssqu.12074

Howard, T. C., & Navarro, O. (2016). Critical Race Theory 20 Years Later: Where Do We Go From Here? *Urban Education*, *51*(3), 253–273. doi: 10.1177/0042085915622541

Hox, J. J. (1993). Factor Analysis of Multilevel Data. Gauging the Muthén Model. In J. Oud & R. van Blokland-Vogelesang (Eds.), *Advances in logitudinal and multivariate analysis in the behavioral sciences* (pp. 141–156). Nijmegen: NL:ITS.

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (Second ed.). New York, NY: Routledge.

Hoxby, C. (2000). *Peer Effects in the Classroom: Learning from Gender and Race Variation.* Retrieved from http://www.usc.es/economet/journals2/eers/eers1134.pdf doi: 10.3386/w7867

Icfes. (2009). *Guía de Orientación Examen de Estado - 2009.* Bogotá.

Icfes. (2010a). *Metodología de Construcción del índice de los Estudiantes -INSE- y de la Clasificación Socieconómica -CSE- de los Colegios* (Tech. Rep.). Bogotá: Author.

Icfes. (2010b). *Orientaciones para el Examen de Estado de la Educación Media ICFES SABER 11°.* Bogotá.

Icfes. (2013). *Guía de acceso a bases de datos Icfes* (Tech. Rep.). Bogotá: Author.

Icfes. (2015). *Lineamientos generales para la presentación del examen de Estado SABER 11 °* (Tech. Rep.). Bogotá: Author.

Icfes. (2017). *Comparability of SABER 11 scores.* Bogotá.

Ilie, S., Sutherland, A., & Vignoles, A. (2017). Revisiting free school meal eligibility as a proxy for pupil socio-economic deprivation. *British Educational Research Journal*, *43*(2), 253–274. Retrieved from `http://doi.wiley.com/10.1002/berj.3260` doi: 10.1002/berj.3260

ILO. (2007). *Pueblos indígenas y afrodescendientes en la región: Hacia la igualdad de oportunidades y el trabajo decente* (Tech. Rep.). Retrieved from `https://www.ilo.org/public/libdoc/ilo/P/09577/09577(2007)38-58.pdf`

Instituto Geográfico Agustín Codazzi - IGAC. (2012). *Mapa Oficial Físico Político-2012.* Bogotá.

Iregui, A. M., Melo, L., & Ramos, J. (2007). Análisis de eficiencia de la educación en Colombia. *Revista de Economía del Rosario*, *10*(1), 21–41. Retrieved from `https://revistas.urosario.edu.co/index.php/economia/article/view/1113`

Iv, W. T., & Hogrebe, M. (2010). School Composition and Context Factors That Moderate and Predict 10th-Grade Science Proficiency. *The Teachers College Record*, *112*(4), 1096–1136. Retrieved from `http://www.tcrecord.org/Content.asp?ContentID=15661`

Ivaniushina, V., & Alexandrov, D. (2018). Anti-school attitudes, school culture and friendship networks. *British Journal of Sociology of Education*, *39*(5), 698–716. Retrieved from `https://www.tandfonline.com/doi/full/10.1080/01425692.2017.1402674` doi: 10.1080/01425692.2017.1402674

Jackson, M. (2012). Bold choices: How ethnic inequalities in educational attainment are suppressed. *Oxford Review of Education*, *38*(2), 189–208. doi: 10.1080/03054985.2012.676249

Jeynes, W. H. (2002). The challenge of controlling for SES in social science and education research. *Educational Psychology Review*, *14*, 205–221. Retrieved from `http://link`

.springer.com/10.1023/A:1014678822410 doi: 10.1023/A:1014678822410

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity*, *52*(4), 1957–1976. Retrieved from http://link.springer.com/10.1007/s11135-017-0584-6 doi: 10.1007/s11135-017-0584-6

Johnston, R., Manley, D., Jones, K., Hoare, A., & Harris, R. (2017). British educational trajectories from school to university: Evaluating quantitative evidence in policy formulation and justification. *Policy and politics*, *45*(2), 137–155. doi: 10.1332/030557316X14598535404750

Jones, J. T., Toma, E. F., & Zimmer, R. W. (2008). School attendance and district and school size. *Economics of Education Review*, *27*(2), 140–148. doi: 10.1016/j.econedurev.2006.09.005

Juhn, C., Murphy, K. M., & Pierce, B. (1993). Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy*, *101*(3), 410–442. Retrieved from https://www.journals.uchicago.edu/doi/10.1086/261881 doi: 10.1086/261881

Kahl, J. A., & Davis, J. A. (1955). A Comparison of Indexes of Socio-Economic Status. *American Sociological Association*, *20*(3), 317–325. Retrieved from http://www.jstor.org/stable/2087391

Kao, G., & Thompson, J. S. (2003). Racial and Ethnic Stratification in Educational Achievement and Attainment. *Annual Review of Sociology*, *29*(2003), 417–442. doi: 10.1146/annurev.soc.29.010202.100019

kei Mak, S., cheung Cheung, K., Soh, K., seong Sit, P., & kai Ieong, M. (2017). An examination of student- and across-level mediation mechanisms accounting for gender differences in reading performance: a multilevel analysis of reading engagement. *Educational Psychology*, *37*(10), 1206–1221. Retrieved from http://dx.doi.org/10.1080/01443410.2016.1242712 doi: 10.1080/01443410.2016.1242712

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, *100*(4), 851–868. doi: 10.1037/0022-0663.100.4.851

Knifsend, C. A., Camacho-Thompson, D. E., Juvonen, J., & Graham, S. (2018). Friends in Activities, School-related Affect, and Academic Outcomes in Diverse Middle

Schools. *Journal of Youth and Adolescence*, *47*(6), 1208–1220. Retrieved from http://dx.doi.org/10.1007/s10964-018-0817-6 doi: 10.1007/s10964-018-0817-6

Konstantopoulos, S., Li, W., Miller, S. R., & van der Ploeg, A. (2017). Do interim assessments reduce the race and SES achievement gaps? *The Journal of Educational Research*, *110*(4), 319–330. Retrieved from http://dx.doi.org/10.1080/00220671.2015.1103685 doi: 10.1080/00220671.2015.1103685

Kotok, S. (2017). Unfulfilled Potential: High-Achieving Minority Students and the High School Achievement Gap in Math. *The High School Journal*, *100*(11), 183–20235. Retrieved from https://doi.org/10.1353/hsj.2017.0007https://muse.jhu.edu/article/649655 doi: 10.1353/hsj.2017.0007

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel Modeling of Individual and Group Level Mediated Effects. *Multivariate Behavioral Research*, *36*(2), 249–277. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S15327906MBR3602_06 doi: 10.1207/S15327906MBR3602_06

Kyriakides, L. (2004). Differential School Effectiveness in Relation to Sex and Social Class: Some Implications for Policy Evaluation. *Educational Research and Evaluation*, *10*(2), 141–161. Retrieved from http://www.tandfonline.com/doi/abs/10.1076/edre.10.2.141.27907 doi: 10.1076/edre.10.2.141.27907

Kyriakides, L., & Creemers, B. P. (2018). Investigating the quality and equity dimensions of educational effectiveness. *Studies in Educational Evaluation*, *57*(December 2017), 1–5. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0191491X17303231 doi: 10.1016/j.stueduc.2017.12.005

Laborda, L., Elosúa, M. R., & Gómez-Veiga, I. (2019). Ethnicity and intelligence in children exposed to poverty environments: An analysis using the Oaxaca-Blinder decomposition. *Intelligence*, *72*, 49–58. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0160289618301156 doi: 10.1016/j.intell.2018.11.008

Lauder, H., Kounali, D., Robinson, T., & Goldstein, H. (2010). Pupil composition and accountability: An analysis in English primary schools. *International Journal of Educational Research*, *49*(2-3), 49–68. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0883035510000492 doi: 10.1016/j.ijer.2010.08.001

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(3), 537–554. Retrieved from http://doi.wiley.com/10.1111/j.1467-985X.2008.00577.x doi: 10.1111/j.1467-985X.2008.00577.x

Leckie, G. (2013a). *Cross-Classified Multilevel Models - Concepts.* Bristol: CMM. Retrieved from http://www.bristol.ac.uk/cmm/learning/course.html

Leckie, G. (2013b). *Multiple Membership Multilevel Models - Concepts.* Bristol: CMM. Retrieved from http://www.bristol.ac.uk/cmm/learning/course.html

Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *172*(4), 835–851. doi: 10.1111/j.1467-985X.2009.00597.x

Leckie, G., & Goldstein, H. (2019). The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal*, 1–20. Retrieved from http://doi.wiley.com/10.1002/berj.3511 doi: 10.1002/berj.3511

Leckie, G., Pillinger, R., Jenkins, J., & Rasbash, J. (2010). School, family, neighbourhood: which is most important to a child's education? *Significance*, *7*(2), 67–70. Retrieved from http://doi.wiley.com/10.1111/j.1740-9713.2010.00421.x doi: 10.1111/j.1740-9713.2010.00421.x

Lee, J. (2002). Racial and Ethnic Achievement Gap Trends: Reversing the Progress Toward Equity? *Educational Researcher*, *31*(1), 3–12. doi: 10.3102/0013189X031001003

Lenkeit, J., Caro, D. H., & Strand, S. (2015). Tackling the remaining attainment gap between students with and without immigrant background: an investigation into the equivalence of SES constructs. *Educational Research and Evaluation*, *21*(1), 60–83. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/13803611.2015.1009915 doi: 10.1080/13803611.2015.1009915

Liu, H., Van Damme, J., Gielen, S., & Van Den Noortgate, W. (2015). School processes mediate school compositional effects: model specification and estimation. *British Educational Research Journal*, *41*(3), 423–447. Retrieved from http://doi.wiley.com/10.1002/berj.3147 doi: 10.1002/berj.3147

Loaiza Quintero, O. L., & Hincapié Vélez, D. (2016). Un estudio de las brechas municipales en calidad educativa en Colombia: 2000-2012. *Ensayos sobre Política Económica*, *34*(79), 3–20. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0120448316000038 doi: 10.1016/j.espe.2016.01.001

Longford, N. T., & Muthén, B. O. (1992). Factor analysis for clustered observations. *Psy-*

chometrika, *57*(4), 581–597. Retrieved from http://link.springer.com/10.1007/BF02294421 doi: 10.1007/BF02294421

Ma, X., Yuan, J., & Luo, X. (2016). Achievement-related within-school socioeconomic gaps in science subjects in China: evidence on existence, consistency, and compensation. *School Effectiveness and School Improvement*, *27*(4), 511–533. Retrieved from http://dx.doi.org/10.1080/09243453.2016.1201122 doi: 10.1080/09243453.2016.1201122

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Journal of Research Methods for the Behavioral and Social Sciences*, *1*(3), 85–91. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc4&AN=2005-10055-001 doi: 10.1027/1614-1881.1.3.86

MacKinnon, D. P. (2008). Multiple Mediator Model. In *Introduction to statistical mediation analysis* (pp. 103–125). New York, NY: Taylor & Francis Group.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Psychology*, *58*(1), 593–614. Retrieved from http://www.annualreviews.org/doi/10.1146/annurev.psych.58.110405.085542 doi: 10.1146/annurev.psych.58.110405.085542

Mannion, G., Sowerby, M., & I 'anson, J. (2015). *How Young People's Participation in School Supports Achievement and Attainment* (Tech. Rep.). Edinburgh: SCCYP.

Manrique, M., Benjumea, S., Rodríguez, I., Nieto, B., Calvo, S. F., Botero, E. S., & Salamanca, M. (2003). *Los pueblos Indígenas en Colombia: derechos, políticas y desafíos*.

Marks, G. N. (2011). Issues in the Conceptualisation and Measurement of Socioeconomic Background: Do Different Measures Generate Different Conclusions? *Social Indicators Research*, *104*(2), 225–251. Retrieved from http://link.springer.com/10.1007/s11205-010-9741-1 doi: 10.1007/s11205-010-9741-1

Marks, H. M. (2002). School composition and peer effects in distinctive organizational settings. *International Journal of Educational Research*, *37*(5), 505–519. doi: 10.1016/S0883-0355(03)00017-X

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom Climate and Contextual Effects: Conceptual and Methodological Issues in the Evaluation of Group-Level Effects. *Educational Psychologist*, *47*(2), 106–124. Retrieved from https://www.tandfonline.com/doi/full/10.1080/00461520.2012.670488 doi: 10.1080/00461520.2012.670488

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). *Doubly-Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error* (Vol. 44) (No. 6). doi: 10.1080/00273170903333665

Marteleto, L., & Andrade, F. (2014). The educational achievement of Brazilian adolescents: Cultural capital and the interaction between families and schools. *Sociology of Education*, *87*, 16–35. Retrieved from http://soe.sagepub.com/content/87/1/16.abstract doi: 10.1177/0038040713494223

Martínez Barrera, A. N. (2012). Un análisis del efecto de pares sobre el rendimiento académico para Colombia. In *Estudios sobre calidad de la educación en colombia* (pp. 99–137). Bogotá: Icfes. Retrieved from http://www.icfes.gov.co/docman/investigadores-y-estudiantes-de-posgrado/resultados-de-investigaciones/factores-asociados/educacion-media-y-basica/1008-un-analisis-del-efecto-pares-sobre-rendimiento-academico-para-colombia-pisa.

Martins, L., & Veiga, P. (2010). Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review*, *29*(6), 1016–1033. Retrieved from http://dx.doi.org/10.1016/j.econedurev.2010.05.001 doi: 10.1016/j.econedurev.2010.05.001

Masel, M. C., Raji, M., & Peek, M. K. (2010). Education and physical activity mediate the relationship between ethnicity and cognitive function in late middle-aged adults. *Ethnicity & Health*, *15*(3), 283–302. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/13557851003681273 doi: 10.1080/13557851003681273

Massey, D. S., & Denton, N. A. (1988). The Dimensions of Residential Segregation. *Social Forces*, *67*(2), 281–315. Retrieved from http://www.jstor.org/stable/2579183?origin=crossref doi: 10.2307/2579183

McEwan, P. J. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of Education Review*, *22*(2), 131–141. doi: 10.1016/S0272-7757(02)00005-5

McEwan, P. J. (2004). The Indigenous Test Score Gap in Bolivia and Chile. *Economic Development and Cultural Change*, *53*(1), 157–190. doi: 10.1086/423257

McEwan, P. J., & Trowbridge, M. (2007). The achievement of indigenous students in Guatemalan primary schools. *International Journal of Educational Development*, *27*(1), 61–76. doi: 10.1016/j.ijedudev.2006.05.004

McKeown, S., Cairns, E., Stringer, M., & Rae, G. (2012). Micro-ecological behavior and intergroup contact. *Journal of Social Psychology*, *152*(3), 340–358. doi: 10.1080/00224545.2011.614647

McKeown, S., & Dixon, J. (2017). The "contact hypothesis": Critical reflections and future directions. *Social and Personality Psychology Compass*, *11*(1), 1–13. doi: 10.1111/spc3.12295

Mickelson, R. A., & Bottia, M. (2009). Integrated education and mathematics outcomes: A synthesis of social science research. *NCL Rev.*, *88*(3), 993–1090.

Mina Calvo, A. (2004). *Factores asociados al logro educativo a nivel municipal.* Bogotá. Retrieved from https://economia.uniandes.edu.co/component/booklibrary/478/view/46/DocumentosCEDE/467/factores-asociados-al-logro-educativo-a-nivel-municipal

Ministry of Culture. (2010). *Afrocolombianos, población con huellas de africanía* (Tech. Rep.). Bogotá: Ministry of Culture. Retrieved from https://www.mincultura.gov.co/areas/poblaciones/comunidades-negras-afrocolombianas-raizales-y-palenqueras/Documents/Caracterizacióncomunidadesnegrasyafrocolombianas.pdf

Ministry of Education. (2003). *Resolution No. 1515: By which the directives, criteria, prodedures and schedule for the organization of the places allocation and registration for preschool, elementary and middle school of public education institutions in local authorities are establish.*

Ministry of Education. (2010). *Mestizaje en Colombia.* Retrieved from http://www.colombiaaprende.edu.co/html/home/1592/article-88867.html

Ministry of Education. (2015a). *Colombia, la mejor educada en el 2025: Líneas estratégicas de la política educativa del Ministerio de Educación Nacional* (Tech. Rep.). Bogotá: Ministry of Education. Retrieved from http://www.mineducacion.gov.co/1621/articles-355154_foto_portada.pdf

Ministry of Education. (2015b). *Decree No. 325 of 2015: By which the Day of Educational Excellence is established.*

Ministry of Education. (2016a). *Buscando colegio.* Bogotá: Ministry of Education. Retrieved from https://sineb.mineducacion.gov.co/bcol/app

Ministry of Education. (2016b). *MEN_ESTUDIANTES-ESTABLECIMIENTOS_EPBM.* Bogotá. Retrieved from https://www.datos.gov.co/Educaci-n/MEN_ESTUDIANTES

-ESTABLECIMIENTOS_EPBM/x4h4-yu8p

Ministry of Education. (2018a). *Atención Educativa a Grupos étnicos.* Bogotá. Retrieved from https://www.mineducacion.gov.co/1759/w3-article-235111.html

Ministry of Education. (2018b). *Matrículas y Pensiones.* Retrieved 2019-05-23, from https://www.mineducacion.gov.co/1759/w3-article-219212.html

Ministry of Inner Affairs. (2014). *Decree No. 1953: By which a special transitory regime for the management of own systems by indigenous communities is created.*

Ministry of Social Protection. (2011). *Resolution No. 3778: by which SISBEN III thresholds are established.* Retrieved from http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=43928

Mohammadpour, E., & Ghafar, M. N. A. (2014). Mathematics achievement as a function of within- and between-school differences. *Scandinavian Journal of Educational Research*, *58*(2), 189–221. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-08873-004 doi: 10.1080/00313831.2012.725097

Mora, J. J., & Arcila, A. M. (2014). Brechas salariales por etnia y ubicación geográfica en Santiago de Cali. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, *18*(1), 34–53. Retrieved from http://www.upo.es/RevMetCuant/pdf/vol18/art95.pdf

Morris, T. T., Davies, N. M., Dorling, D., Richmond, R. C., & Smith, G. D. (2018). Testing the validity of value-added measures of educational progress with genetic data. *British Educational Research Journal*, *44*(5), 725–747. doi: 10.1002/berj.3466

Morris, T. T., Manley, D., & Van Ham, M. (2018). Context or composition: How does neighbourhood deprivation impact upon adolescent smoking behaviour? *PLoS ONE*, *13*(2), 1–16. doi: 10.1371/journal.pone.0192566

Moullin, S., Waldfogel, J., & Washbrook, E. (2018). Parent-child attachment as a mechanism of intergenerational (dis)advantage. *Families, Relationships and Societies*, *7*(2), 265–284. Retrieved from http://www.ingentaconnect.com/content/10.1332/204674317X15071998786492 doi: 10.1332/204674317X15071998786492

Muñoz Wilches, L. A. (2010). Estudios de factores asociados a la calidad de la educación escolar de Bogotá. *Euducación y Ciudad*(19), 58–68.

Muller, C., Riegle-Crumb, C., Schiller, K. S., Wilkinson, L., & Frank, K. a. (2010). Race and Academic Achievement in Racially Diverse High Schools: Opportunity and

Stratification. *Teachers College Record*, *112*(4), 1038–1063. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2893342/

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, *46*(1), 69–85. Retrieved from http://www.jstor.org/stable/1913646

Munk, T. E., McMillian, M. M., & Lewis, N. R. (2014). Compositional Effects, Segregation and Test Scores: Evidence From the National Assessment of Educational Progress. *The Review of Black Political Economy*, *41*(4), 433–454. Retrieved from http://link.springer.com/10.1007/s12114-014-9200-3 doi: 10.1007/s12114-014-9200-3

Muñoz-Chereau, B. (2018). Exploring gender gap and school differential effects in mathematics in Chilean primary schools. *School Effectiveness and School Improvement*, *00*(00), 1–21. Retrieved from https://doi.org/10.1080/09243453.2018.1503604 doi: 10.1080/09243453.2018.1503604

Murillo, F. J. (2016). Midiendo la segregación escolar en América Latina. Un análisis metodológico utilizando el TERCE / Measuring the School Segregation in Latin America. A Methodological Analysis using TERCE. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *14*(4), 33–60. Retrieved from https://revistas.uam.es/index.php/reice/article/view/6496 doi: 10.15366/reice2016.14.4.002

Murphy, J. (2009). Closing achievement gaps: Lessons from the last 15 years. *Phi Delta Kappa International*, *91*(3), 8–12.

Myers, S. L., Kim, H., & Mandala, C. (2004). The effect of school poverty on racial gaps in test scores: The case of the Minnesota basic standards tests. *The Journal of Negro Education*, *73*, 81. Retrieved from http://www.jstor.org/stable/10.2307/3211261?origin=crossref doi: 10.2307/3211261

Naciones Unidas. (2009). *La Situación De Los Pueblos Indígenas En Colombia: Seguimento a Las Recomendaciones Hechas Por El Relator Especial Anterior* (Tech. Rep.). Washington, D.C.: United Nations.

Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: the Big-Fish-Little-Pond-Effect for PISA 2006. *Educational Psychology*, *31*(5), 629–656. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/01443410.2011.586416 doi: 10.1080/01443410.2011.586416

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology*, *104*(4), 1033–1053. doi: 10.1037/a0027697

National Center for Education Statistics. (2016). *International Activities Program - IAP.* Retrieved from https://nces.ed.gov/surveys/SurveyGroups.asp?group=6

Navarro Piedrahita, E. G., & Tovar, N. A. (2008). *Cambio de Calendario Escolar iniciará a partir del mes de Agosto.* Cali: Secretaría de Educación del Valle. Retrieved from http://www.valledelcauca.gov.co/educacion/publicaciones.php?id=5592

Ndum, E., Allen, J., Way, J., & Casillas, A. (2018). Explaining Gender Gaps in English Composition and College Algebra in College: The Mediating Role of Psychosocial Factors. *Journal of Advanced Academics*, *29*(1), 56–88. Retrieved from http://journals.sagepub.com/doi/10.1177/1932202X17740331 doi: 10.1177/1932202X17740331

Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between-and Within-Cluster Covariate Effects in the Analysis of Clustered Data Effects in the Analysis of Clustered Data. *Biometrics*, *54*(2), 638–645. Retrieved from http://www.jstor.org/stable/3109770

Newsom, J. (1963). *Half Our Future* (Tech. Rep.). London: Central Advisory Council for Education (England). Retrieved from http://www.educationengland.org.uk/documents/newsom/newsom1963.html

Nicol, C., Caplen, G., Statham, P., & Browne, W. (2011). Decisions about foraging and risk trade-offs in chickens are associated with individual somatic response profiles. *Animal Behaviour*, *82*(2), 255–262. Retrieved from http://dx.doi.org/10.1016/j.anbehav.2011.04.022https://linkinghub.elsevier.com/retrieve/pii/S0003347211001837 doi: 10.1016/j.anbehav.2011.04.022

Noe, D., Rodriguez Cabello, J., & Zuñiga, I. (2005). *Brecha étnica e influencia de los pares en el rendimiento escolar: Evidencia para Chile*. Santiago de Chile: CEPAL. Retrieved from http://www.cepal.org/es/publicaciones/6092-brecha-etnica-e-influencia-de-los-pares-en-el-rendimiento-escolar-evidencia-para

Núñez, J., Steiner, R., Cadena, X., & Pardo, R. (2002). *¿Cuáles colegios ofrecen mejor educación en Colombia?* Bogotá. Retrieved from https://economia.uniandes.edu.co/components/com_booklibrary/ebooks/D2002-03.pdf

Nuttall, D. L., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, *13*(7), 769–776. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/088303558990027X doi: 10.1016/0883-0355(89)90027-X

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*(3), 693. Retrieved from http://www.jstor.org/stable/2525981?origin=crossref doi: 10.2307/2525981

OECD. (2016). *Education in Colombia*. Paris: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/education-in-colombia_9789264250604-en doi: 10.1787/9789264250604-en

Ogbu, J. U. (1995a). Cultural problems in minority education: Their interpretations and consequences—Part one: Theoretical Background. *Urban Review*, *27*(4), 271. Retrieved from http://0-search.ebscohost.com.libraries.colorado.edu/login.aspx?direct=true&db=aph&AN=9601112284&site=ehost-live&scope=site&scope=cite

Ogbu, J. U. (1995b). Cultural problems in minority education: Their interpretations and consequences—Part two: Case studies. *The Urban Review*, *27*(4), 271–297. Retrieved from http://link.springer.com/10.1007/BF02354409 doi: 10.1007/BF02354409

Ojeda Rosero, D. E., & González Gómez, M. P. (2012). Percepciones y estereotipos de estudiantes universitarios hacia compañeros afrocolombianos e indígenas. *Revista CES Psicología*, *5*(2), 2011–3080.

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives*, *25*(1), 159–184. Retrieved from http://pubs.aeaweb.org/doi/10.1257/jep.25.1.159 doi: 10.1257/jep.25.1.159

Orjuela, J. (2013). Análisis del desempeño estudiantil en las pruebas de Estado para educación media en Colombia mediante modelos jerárquicos lineales. *Ingeniería*, *18*(2), 54–67. Retrieved from http://revistas.udistrital.edu.co/ojs/index.php/reving/article/view/5231

Osborne, J. W. (2001). Testing Stereotype Threat: Does Anxiety Explain Race and Sex Differences in Achievement? *Contemporary Educational Psychology*, *26*(3), 291–310. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0361476X0091052X doi: 10.1006/ceps.2000.1052

Østby, G., & Urdal, H. (2010). *Education and Civil Conflict: A Review of the Quantitative,* *Empirical Literature* (Tech. Rep.). Oslo: UNESCO.

Page, L., Murnane, R., & Willett, J. (2008). *Trends in the Black-White Achievement Gap: Clarifying the Meaning of Within- and Between-School Achievement Gaps.* Cambridge, MA. Retrieved from `http://www.nber.org/papers/w14213.pdf` doi: 10.3386/w14213

Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: a multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, *19*(1), 21–49. Retrieved from `http://www.tandfonline.com/doi/abs/10.1080/09243450801936845` doi: 10.1080/09243450801936845

Palardy, G. J. (2015). High school socioeconomic composition and college choice: multilevel mediation via organizational habitus, school practices, peer and staff attitudes. *School Effectiveness and School Improvement*, *26*(3), 329–353. Retrieved from `http://dx.doi.org/10.1080/09243453.2014.965182` doi: 10.1080/09243453.2014.965182

Palardy, G. J., Rumberger, R. W., & Butler, T. (2015). The effect of high school socioeconomic, racial, and linguistic segregation on academic performance and school behaviors. *Teachers College Record*, *117*(December).

Patacchini, E., & Zenou, Y. (2016). Racial identity and education in social networks. *Social Networks*, *44*, 85–94. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S0378873315000507` doi: 10.1016/j.socnet.2015.06.001

Payne, M. R. (2003). The differential impact of family characteristics on the academic achievement of black and white youth. *Race and Society*, *6*(2), 141–162. doi: 10.1016/j.racsoc.2004.11.006

Pérez Marulanda, L., & Mora, J. J. (2014). La calidad del empleo en la población afrodescendiente colombiana: una aproximación desde la ubicación geográfica de las comunas (Job Quality in the Colombian Afro-Descendant Population: An Approach from the Geographic Location of the City Districts). *Revista de Economía del Rosario*, *17*(24), 315–347. Retrieved from `http://revistas.urosario.edu.co/index.php/economia/article/view/3748/2702` doi: 10.12804/rev.econ.rosario.17.02.2014.05

Perry, T. (2016). English Value-Added Measures: Examining the Limitations of School Performance Measurement. *British Educational Research Journal*, *42*(6), 1056–1080. doi: 10.1002/berj.3247

Phoenix, A., & Husain, F. (2007). *Parenting and Ethnicity* (Tech. Rep.). York: Joseph Rowntree Foundation.

Piñeros Jiménez, L. J., & Rodríguez, A. (1998). Los Insumos Escolares en la Educación Secundaria y su Efecto Sobre el Rendimiento Académico de los Estudiantes: Un estudio en Colombia. *The World Bank Paper Series*, *36*(December), 5–60. Retrieved from http://datatopics.worldbank.org/hnp/files/edstats/COLlac98.pdf

Pituch, K. A., & Stapleton, L. M. (2012). Distinguishing Between Cross- and Cluster-Level Mediation Processes in the Cluster Randomized Trial. *Sociological Methods and Research*, *41*(4), 630–670. doi: 10.1177/0049124112460380

Plewis, I. (2011). Contextual variations in ethnic group differences in educational attainments. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*, 419–437. Retrieved from http://doi.wiley.com/10.1111/j.1467-985X.2010.00679.x doi: 10.1111/j.1467-985X.2010.00679.x

Plomin, R. (2018). *Blueprint: How DNA makes us who we are*. Allen Lane.

Plowden, B. H. (1967). Children of Immigrants. In *Children and their primary schools* (pp. 69–74). London: Her Majesty's Stationery Office. Retrieved from http://www.educationengland.org.uk/documents/plowden/plowden1967-1.html#06

Pokropek, A., Borgonovi, F., & Jakubowski, M. (2015). Socio-economic disparities in academic achievement: A comparative analysis of mechanisms and pathways. *Learning and Individual Differences*, *42*(2015), 10–18. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S1041608015001478 doi: 10.1016/j.lindif.2015.07.011

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10.1037/a0020141 doi: 10.1037/a0020141

Presidency of the Republic of Colombia. (1994). *Dercree 1860 of 1994: By which the Law 115 of 1994 is partially regulated regarding the general pedagogical and organizational aspects.* Retrieved from http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=1289

Presidency of the Republic of Colombia. (2002a). *Decree 1850 of 2002: By which school day is regulated.* Retrieved from http://www.mineducacion.gov.co/1621/article-103274.html

Presidency of the Republic of Colombia. (2002b). *Decree 230 of 2002: By which the norms regarding curriculum, evaluation and promotion of students and institutional evaluation are dictated.* Retrieved from http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=4684

Presidency of the Republic of Colombia. (2002c). *Decree No. 1278 of 2002: By which the teacher professionalization statute is created.* doi: -ISBN978-92-75-32913-9

Presidency of the Republic of Colombia. (2011). *Decree 4807: By which the conditions for the application of free education for pre-school, elementary and middle education of public institutions are established and other dispositions regarding its implementation are dictated.*

Presidency of the Republic of Colombia. (2012a). *Decree No 1231 of 2012: By which the Decree No 829 of 2012 is modified.*

Presidency of the Republic of Colombia. (2012b). *Decree No 827 of 2012: By which the teachers' payments are modified.*

Quinn, D. M. (2015a). Black-White Summer Learning Gaps: Interpreting the Variability of Estimates Across Representations. *Educational Evaluation and Policy Analysis*, *37*(1), 50–69. Retrieved from http://epa.sagepub.com/cgi/doi/10.3102/0162373714534522 doi: 10.3102/0162373714534522

Quinn, D. M. (2015b). Kindergarten Black–White Test Score Gaps. *Sociology of Education*, *88*(2), 120–139. Retrieved from http://journals.sagepub.com/doi/10.1177/0038040715573027 doi: 10.1177/0038040715573027

Quinn, D. M., & Le, Q. T. (2018). Are We Trending to More or Less Between-Group Achievement Inequality Over the School Year and Summer? Comparing Across ECLS-K Cohorts. *AERA Open*, *4*(4), 1–19. Retrieved from http://journals.sagepub.com/doi/10.1177/2332858418819995 doi: 10.1177/2332858418819995

Quintero Ramírez, O. (2017). Aprendiendo la "R": racialización y racismo prosaico en escuelas bogotanas (Learning the "R": Racialisation and Prosaic Racism within Schools in Bogota). *Revista Antropologías del Sur*, 105–125.

R Core Team. (2017). *R: A Language and Environment for Statistical Computing.* Vienna.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata* (Third Edit ed.). College Station: Stata Press. Retrieved from http://www.stata.com/bookstore/multilevel-longitudinal-modeling-stata/

Ramírez, L., Urueña, H., & Moreno, P. (2012). Discursos esencialistas sobre la raza en Colombia y algunas implicaciones sociales (Essentialist discourse about race in Colombia and social implications). *Revista Latinoamericana de Psicología*, *44*(3), 119–131.

Rangel, C., & Lleras, C. (2010). Educational inequality in Colombia: Family background, school quality and student achievement in Cartagena. *International Studies in Sociology of Education*, *20*, 291–317. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/09620214.2010.530855 doi: 10.1080/09620214.2010.530855

Rangel, M. (2001). *La inequidad étnico-racial y la formación para el trabajo en América Latina y el Caribe* (Tech. Rep.). Santiago de Chile: CEPAL and GTZ.

Rasbash, J., Leckie, G., Pillinger, R., & Jenkins, J. (2010). Children's educational progress: partitioning family, school and area effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(3), 657–682. Retrieved from http://doi.wiley.com/10.1111/j.1467-985X.2010.00642.x doi: 10.1111/j.1467-985X.2010.00642.x

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models Applications and Data Analysis Methods* (Second Edi ed.). Thousand Oaks: SAGE Publishing. Retrieved from https://uk.sagepub.com/en-gb/eur/hierarchical-linear-models/book9230

Ready, D. D., & Reid, J. L. (2018). Children's executive function development and school socio-economic and racial/ethnic composition. *Early Childhood Research Quarterly*. Retrieved from https://doi.org/10.1016/j.ecresq.2018.08.002https://linkinghub.elsevier.com/retrieve/pii/S0885200618300863 doi: 10.1016/j.ecresq.2018.08.002

Reardon, S. F. (2008). *Thirteen Ways of Looking at the Black-White Test Score Gap.* Standford. Retrieved from http://web.stanford.edu/group/cepa/workingpapers/WORKING_PAPER_2008_08.pdf

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White Achievement Gap in Math and Reading in the Elementary Grades. *American Educational Research Journal*, *46*(3), 853–891. Retrieved from http://aer.sagepub.com/cgi/doi/10.3102/0002831209333184 doi: 10.3102/0002831209333184

Reardon, S. F., & Ho, A. D. (2015). Practical Issues in Estimating Achievement Gaps From Coarsened Data. *Journal of Educational and Behavioral Statistics*, *40*(2), 158–189. Retrieved from http://jeb.sagepub.com/cgi/doi/10.3102/1076998615570944 doi:

10.3102/1076998615570944

Red Nacional de Información. (2018). *Cifras Unidad de Víctimas.* Retrieved 2018-11-04, from https://cifras.unidadvictimas.gov.co/

Redacción de El País. (2010, oct). *Colegios inician cambio del calendario B al A.* Cali. Retrieved from http://www.elpais.com.co/elpais/cali/noticias/colegios-inician-cambio-del-calendario-b

Redacción El Espectador. (2010, aug). *Educación , gratis para todos.* Bogotá. Retrieved from https://www.elespectador.com/publicaciones/especial/articulo-217522-educacion-gratis-todos

Rentería, C. (2009). *Avances y retos de la Política Social en Colombia* (Tech. Rep.). Bogotá: DNP. Retrieved from https://colaboracion.dnp.gov.co/CDT/Prensa/Publicaciones/15-AvancesdelaPolíticaSocialenColombia.pdf

Rich, Y., Ari, R. B., Amir, Y., & Eliassy, L. (1996, jun). Effectiveness of schools with a mixed student body of natives and immigrants. *International Journal of Intercultural Relations*, *20*(3-4), 323–339. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/0147176796000223 doi: 10.1016/0147-1767(96)00022-3

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, *73*(2), 417–458. Retrieved from http://doi.wiley.com/10.1111/j.1468-0262.2005.00584.x doi: 10.1111/j.1468-0262.2005.00584.x

Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E., & Stanat, P. (2014). Socioeconomic and language minority classroom composition and individual reading achievement: The mediating role of instructional quality. *Learning and Instruction*, *32*, 63–72. Retrieved from http://dx.doi.org/10.1016/j.learninstruc.2014.01.007 doi: 10.1016/j.learninstruc.2014.01.007

Robinson, J. P. (2008). Evidence of a Differential Effect of Ability Grouping on the Reading Achievement Growth of Language-Minority Hispanics. *Educational Evaluation and Policy Analysis*, *30*(2), 141–180. Retrieved from http://journals.sagepub.com/doi/10.3102/0162373708317742 doi: 10.3102/0162373708317742

Rodríguez, C., & Sánchez, F. (2012). Armed Conflict Exposure, Human Capital Investments, And Child Labor: Evidence From Colombia. *Defence and Peace Economics*, *23*(2), 161–184. doi: 10.1080/10242694.2011.597239

Rojas Curieux, T. (2010). La etnoeducación en Colombia: un trecho andado y un largo camino por recorrer. *Colombia Internacional*(46), 45–59. Retrieved from https://colombiainternacional.uniandes.edu.co/view.php/349/index.php?id=349

Romero Mora, B., Genes Díaz, J., & Ospino Meriño, I. (2017). Mercado laboral informal en las ciudadades de Riohacha y Maicao, La guajira – Colombia. *Redipe*, *6*(6), 132 –143. Retrieved from http://revista.redipe.org/index.php/1/article/view/251

Rosenbaum, J. E., & DeLuca, S. (2008). What Kinds of Neighborhoods Change Lives? The Chicago Gautreaux Housing Program and Recent Mobility Programs. *Indiana Law Review*, *41*(3), 653–662. Retrieved from https://mckinneylaw.iu.edu/ilr/pdf/vol41p653.pdf

Rothstein, R. (2004). *Class And Schools: Using Social, Economic, And Educational Reform To Close The Black-White Achievement Gap*. New York, NY: Teachers College Press.

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*(1), 127–141. doi: 10.1002/sim.2331

Rozo Alzate, J. (2017). La educación secundaria y sus dos dimensiones. Efectos del barrio y del colegio sobre los resultados saber 11. *Revista de Economía del Rosario*, *20*(1), 38. Retrieved from https://revistas.urosario.edu.co/index.php/economia/article/view/6148 doi: 10.12804/revistas.urosario.edu.co/economia/a.6148

Rutkowski, D., & Rutkowski, L. (2013). Measuring socioeconomic background in PISA: One size might not fit all. *Research in Comparative and International Education*, *8*, 259. Retrieved from http://rci.sagepub.com/lookup/doi/10.2304/rcie.2013.8.3.259 doi: 10.2304/rcie.2013.8.3.259

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (1st ed., Vol. 3, pp. 249–277). Amsterdam: Elsevier B.V. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/B9780444534293000041 doi: 10.1016/B978-0-444-53429-3.00004-1

Sammons, P. (2007). *School Effectiveness and Equity: Making Connections. A review of school effectiveness and improvement research - its implications for practitioners and policy makers* (Tech. Rep. No. July). Retrieved from www.cfbt.com/evidenceforeducation/pdf/FullLiteratureReview.pdf

Sánchez-Jabba, A. (2011). Ethnic background and academic performance in Colombia. *Revista de Economía del Rosario*, *14*(2), 189–227.

Santibañez, L. (2016). The indigenous achievement gap in Mexico: The role of teacher policy under intercultural bilingual education. *International Journal of Educational Development*, *47*, 63–75. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0738059315300183 doi: 10.1016/j.ijedudev.2015.11.015

Schindler Rangvid, B. (2007). School composition effects in Denmark: quantile regression evidence from PISA 2000. *Empirical Economics*, *33*(2), 359–388. Retrieved from http://link.springer.com/10.1007/s00181-007-0133-6 doi: 10.1007/s00181-007-0133-6

Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The Role of Schooling in Perpetuating Educational Inequality. *Educational Researcher*, *44*(7), 371–386. Retrieved from http://edr.sagepub.com/cgi/doi/10.3102/0013189X15603982 doi: 10.3102/0013189X15603982

Schmidt, W. H., Zoido, P., & Cogan, L. S. (2013). *Schooling matters: Opportunity to learn in PISA 2012.* doi: 10.1787/5k3v0hldmchl-en

Secretaría de Educación Municipal de Santiago de Cali. (2008). *Resolution No. 4143.2.21.4739 of 2008, by which the 2008-2009 school year is modified.*

Shunck, R. (2010). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *Stata Journal*, *10*(3), 288–308. Retrieved from http://ideas.repec.org/a/tsj/stataj/v7y2007i4p465-506.html doi: TheStataJournal

Simonite, V. (2005). Gender difference added? Institutional variations in the gender gap in first class degree awards in mathematical sciences. *British Educational Research Journal*, *31*(6), 737–759. doi: 10.1080/01411920500314885

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417–453. Retrieved from http://rer.sagepub.com/cgi/doi/10.3102/00346543075003417 doi: 10.3102/00346543075003417

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563–575. doi: 10.1007/BF02296196

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Second ed.). London: SAGE.

Sohn, K. (2012). The dynamics of the evolution of the Black–White test score gap. *Education Economics*, *20*(2), 175–188. doi: 10.1080/09645292.2010.488487

Sonnenschein, S., & Galindo, C. (2014). Race/Ethnicity and Early Mathematics Skills: Relations Between Home, Classroom, and Mathematics Achievement. *The Journal of Educational Research*, *0671*(November), 1–17. Retrieved from `http://www.tandfonline.com/doi/full/10.1080/00220671.2014.880394` doi: 10.1080/00220671.2014.880394

Soto Ceballos, M. (2014). *La Desigualdad en el Logro Educativo: Un análisis para la Región del Pacífico* (Undergrad). Universidad del Valle.

StataCorp. (2017). *Stata Statistical Software: Release 15.* College Station, TX.

Stein, A., Behrman, J., DiGirolamo, A., Grajeda, R., Martorell, R., Quisumbing, A., & Ramakrishnan, U. (2005). Schooling, educational achievement, and cognitive functioning among young guatemalan adults. *Food and Nutrition Bulletin*, *26*(2 SUPPL. 1), S46-S54. Retrieved from `https://www.scopus.com/inward/record.uri?eid=2-s2.0-21344460473&partnerID=40&md5=75f42f14bfca5f7c51bae66b4b9e9d5d` (cited By 21)

Stevens, P. A. J., & Dworkin, A. G. (Eds.). (2014). *The Palgrave Handbook of Race and Ethnic Inequalities in Education*. London: Palgrave Macmillan UK. Retrieved from `http://link.springer.com/10.1057/9781137317803` doi: 10.1057/9781137317803

Stevens, T., Olivárez, A., & Hamman, D. (2006, may). The Role of Cognition, Motivation, and Emotion in Explaining the Mathematics Achievement Gap Between Hispanic and White Students. *Hispanic Journal of Behavioral Sciences*, *28*(2), 161–186. Retrieved from `http://journals.sagepub.com/doi/10.1177/0739986305286103` doi: 10.1177/0739986305286103

Stiefel, L., Schwartz, A. E., & Gould Ellen, I. (2007). Disentangling the racial test score gap: Probing the evidence in a large urban school district. *Journal of Policy Analysis and Management*, *26*(1), 7–30. Retrieved from `http://doi.wiley.com/10.1002/pam.20225` doi: 10.1002/pam.20225

Stokes, L., Rolfe, H., Hudson-Sharp, N., Stevens, S., & Strand, S. (2015). *A compendium of evidence on ethnic minority resilience to the effects of deprivation on attainment Research report* (No. DFE-RR439). London: National Institute of Economic and Social Research & University of Oxford. Retrieved from `https://www.gov.uk/government/publications/ethnic-minorities-and-attainment-the-effects-of-poverty`

Strand, S. (1999). Ethnic group, sex and economic disadvantage: Associations with pupils' educational progress from Baseline to the end of Key Stage 1. *British Educational Research Journal*, *25*(2), 179–202. doi: 10.1080/0141192990250204

Strand, S. (2010, sep). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement*, *21*(3), 289–314. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/09243451003732651 doi: 10.1080/09243451003732651

Strand, S. (2011, apr). The limits of social class in explaining ethnic gaps in educational attainment. *British Educational Research Journal*, *37*(2), 197–229. Retrieved from http://doi.wiley.com/10.1080/01411920903540664 doi: 10.1080/01411920903540664

Strand, S. (2012). The White British-Black Caribbean achievement gap: Tests, tiers and teacher expectations. *British Educational Research Journal*, *38*(1), 75–101. doi: 10.1080/01411926.2010.526702

Strand, S. (2014a). Ethnicity, gender, social class and achievement gaps at age 16: intersectionality and 'getting it' for the white working class. *Research Papers in Education*, *29*(2), 131–171. Retrieved from http://dx.doi.org/10.1080/02671522.2013.767370 doi: 10.1080/02671522.2013.767370

Strand, S. (2014b). School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11. *Oxford Review of Education*, *40*, 223–245. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/03054985.2014.891980 doi: 10.1080/03054985.2014.891980

Strand, S. (2016). Do some schools narrow the gap? Differential school effectiveness revisited. *Review of Education*, *4*(2), 107–144. Retrieved from http://doi.wiley.com/10.1002/rev3.3054 doi: 10.1002/rev3.3054

Sussman, R. W. (2016). *The Myth of Race: the Troubling Persistence of an Unscientific Idea*. Cambridge: Harvard University Press.

Sutherland, A., Ilie, S., & Vignoles, A. (2015). *Factors associated with achievement: Key stage 4* (Tech. Rep. No. November). Department for Education. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/473673/RR407_-_Factors_associated_with_achievement_-_key_stage_4.pdf

Swann, L. (1985). *Education for All* (Tech. Rep.). London: Committee of Enquiry into the Education of Children from Ethnic Minority Groups. Retrieved from http://

www.educationengland.org.uk/documents/swann/swann1985.html

Taylor, M. J. (1981). *Caught between: A review of research into the education of pupils of West Indian origin* (Tech. Rep.). London: Committee of Inquiry into the Education of Children from Ethnic Minority Groups.

Templer, D. I., & Arikawa, H. (2006). Temperature, skin color, per capita income, and IQ: An international perspective. *Intelligence*, *34*(2), 121–139. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0160289605000917 doi: 10.1016/j.intell.2005.04.002

The World Bank. (2016). *Colombia.* Retrieved from http://data.worldbank.org/country/colombia

Thomas, M. K. (2004). The SAT II: Minority/majority test-score gaps and what they could mean for college admissions. *Social Science Quarterly*, *85*(5 SPEC. ISS.), 1318–1334. doi: 10.1111/j.0038-4941.2004.00278.x

Thomas, S., Salim, M. M., Muñoz-Chereau, B., & Peng, W.-J. (2012). Educational quality, effectiveness and evaluation: Perspectives from China, South America and Africa. In C. Chapman (Ed.), *School effectiveness and improvement research, policy and practice: Challenging the orthodoxy?* (pp. 125–146). Abingdon: Routledge. Retrieved from https://www.routledge.com/School-Effectiveness-and-Improvement-Research-Policy-and-Practice-Challenging/Chapman-Armstrong-Harris-Muijs-Reynolds-Sammons/p/book/9780415698993

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997, sep). Differential Secondary School Effectiveness: comparing the performance of different pupil groups. *British Educational Research Journal*, *23*(4), 451–469. Retrieved from http://doi.wiley.com/10.1080/0141192970230405 doi: 10.1080/0141192970230405

Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, *37*(5), 483–504. doi: 10.1016/S0883-0355(03)00016-8

Todd, P. E., & Wolpin, K. I. (2007, dec). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, *1*(1), 91–136. Retrieved from http://www.journals.uchicago.edu/doi/10.1086/526401 doi: 10.1086/526401

Tofighi, D., MacKinnon, D. P., & Yoon, M. (2009). Covariances between regression coefficient estimates in a single mediator model. *British Journal of Mathematical* *and Statistical Psychology*, *62*(3), 457–484. Retrieved from http://doi.wiley.com/10.1348/000711008X331024 doi: 10.1348/000711008X331024

Tofighi, D., & Thoemmes, F. (2014). Single-Level and Multilevel Mediation Analysis. *The Journal of Early Adolescence*, *34*(1), 93–119. Retrieved from http://journals.sagepub.com/doi/10.1177/0272431613511331 doi: 10.1177/0272431613511331

Turner, H., Rubie-Davies, C. M., & Webber, M. (2015). Teacher Expectations, Ethnicity and the Achievement Gap. *New Zealand Journal of Educational Studies*, *50*(1), 55–69. Retrieved from http://link.springer.com/10.1007/s40841-015-0004-1 doi: 10.1007/s40841-015-0004-1

Ugarriza, J. E., & Craig, M. J. (2012). *The Relevance of Ideology to Contemporary Armed Conflicts* (Vol. 57) (No. 3). doi: 10.1177/0022002712446131

UNESCO Institute for Statistics. (2012). *Opportunities lost: The impact of grade repetition and early school leaving*. Montreal: UNESCO. Retrieved from http://learningportal.iiep.unesco.org/en/notice/032164

UNESCO-OREALC. (2016). *Reporte Técnico. Tercer Estudio Regional Comparativo y Explicativo. TERCE*. Santiago de Chile: Oficina Regional de Educación para América Latina y el Caribe (OREALC/UNESCO Santiago).

United Nations Development Programme. (2015). *Multidimensional Poverty Index (MPI).* Retrieved from http://hdr.undp.org/en/content/multidimensional-poverty-index-mpi

University of Bristol. (2016). *University of Bristol Research Governance and Integrity Policy* (No. October). Bristol: University of Bristol. Retrieved from http://www.bristol.ac.uk/media-library/sites/red/documents/research-governance/rgi.pdf

Valoyes Chavez, L. E. (2015). "Blacks aren't good for math": Racial ideologies and mathematics teaching practices in Colombia. *CS*, *16*, 169–206. doi: 10.18046/recs.i15.1909

Valoyes-Chávez, L., & Martin, D. B. (2016). Exploring racism inside and outside the mathematics classroom in two different contexts: Colombia and USA. *Intercultural Education*, *27*(4), 363–376. Retrieved from http://dx.doi.org/10.1080/14675986.2015.1106135 doi: 10.1080/14675986.2015.1106135

van Ewijk, R., & Sleegers, P. (2010a). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, *5*(2), 134–150. Retrieved from http://dx.doi.org/10.1016/j.edurev.2010.02.001 doi: 10.1016/j.edurev

.2010.02.001

van Ewijk, R., & Sleegers, P. (2010b). Peer ethnicity and achievement: a meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, *21*(3), 237–265. doi: 10.1080/09243451003612671

van Tubergen, F., & van Gaans, M. (2016). Is There an Oppositional Culture Among Immigrant Adolescents in the Netherlands? *Youth & Society*, *48*(2), 202–219. Retrieved from `http://journals.sagepub.com/doi/10.1177/0044118X13486901` doi: 10.1177/0044118X13486901

Velez Bustillo, E., & Psacharopoulos, G. (2019). *Colombia's "Ser Pilo Paga" program: A smart investment?* Retrieved 2019-02-09, from `http://blogs.worldbank.org/education/colombia-s-ser-pilo-paga-program-smart-investment`

Verhaeghe, J. P., Vanlaar, G., Knipprath, H., De Fraine, B., & Van Damme, J. (2018). Can group composition effects explain socioeconomic and ethnic achievement gaps in primary education? *Studies in Educational Evaluation*, *57*(July 2017), 6–15. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S0191491X1630147X` doi: 10.1016/j.stueduc.2017.07.006

Vertel Morinson, M. L., Cepeda Coronado, J. A., & Lugo Hernández, E. A. (2014). Análisis Multivariado de la Calidad educativa en Sucre. *Scientia et Technica*, *19*(1), 96 – 105. Retrieved from `http://revistas.utp.edu.co/index.php/revistaciencia/article/view/8953`

Viáfara López, C. A., & Serna Alvarado, N. J. (2015). Desigualdad de oportunidades educativas en la población de 15 a 29 años en Brasil y Colombia según autoclasificación étnico-racial (Inequality of Educational Opportunities in the 20 to 29 Years Old Population of Brazil and Colombia by Ethnic-Racial Self-A. *Sociedad y Economía*, *2015*(29), 151–174. Retrieved from `http://www.redalyc.org/articulo.oa?id=99639245009`

Villegas García, M., & López Quiroz, L. (2011). Apartheid educativo: Educacion, desigualdad e inmovilidad social en Bogota. *Revista de Economia Institucional*, *13*(25), 137–162.

Vinha, L. G. d. A., Karino, C. A., & Laros, J. A. (2016). Factors Associated with Mathematics Performance in Brazilian Basic Education. *Psico-USF*, *21*(1), 87–100. Retrieved from `http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-82712016000100087&lng=en&nrm=iso&tlng=en` doi: 10

.1590/1413-82712016210108

Voight, A., Hanson, T., O'Malley, M., & Adekanye, L. (2015). The Racial School Climate Gap: Within-School Disparities in Students' Experiences of Safety, Support, and Connectedness. *American Journal of Community Psychology*, *56*(3-4), 252–267. Retrieved from http://doi.wiley.com/10.1007/s10464-015-9751-x doi: 10.1007/s10464-015-9751-x

Wenglinsky, H. (2004). Closing the racial achievement gap: The role of reforming instructional practices. *Education Policy Analysis Archives*, *12*, 1–22.

Wenglinsky, H. (2009). How Money Matters : The Effect of School District Spending on Academic Achievement. *Sociology The Journal Of The British Sociological Association*, *70*(3), 221–237.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*, 461–481. Retrieved from http://content.apa.org/journals/bul/91/3/461 doi: 10.1037/0033-2909.91.3.461

Whitt, H. P. (1986). The sheaf coefficient: A simplified and expanded approach. *Social Science Research*, *15*(2), 174–189. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/0049089X86900141 doi: 10.1016/0049-089X(86)90014-1

Wicherts, J. M., Borsboom, D., & Dolan, C. V. (2010). Why national IQs do not support evolutionary theories of intelligence. *Personality and Individual Differences*, *48*(2), 91–96. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0191886909002475 doi: 10.1016/j.paid.2009.05.028

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. Retrieved from http://ggplot2.org

Wiggan, G. (2007). Race, School Achievement, and Educational Inequality: Toward a Student-Based Inquiry Perspective. *Review of Educational Research*, *77*(3), 310–333. doi: 10.3102/003465430303947

Wilkinson, I. A. G., & Fung, I. Y. Y. (2002). Small-group composition and peer effects. *International Journal of Educational Research*, *37*(5), 425–447. doi: 10.1016/S0883-0355(03)00014-4

Williams, J. H. (2004). Civil Conflict, Education, and the Work of Schools: Twelve Propositions. *Conflict Resolution Quarterly*, *21*(4), 471–481. doi: 10.1002/crq.75

Willke, I. (1975). Schooling of immigrant children in West-Germany - Sweden - England: The educationally disadvantaged. *International Review of Education*. doi: 10.1007/

BF00615656

Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. London: The Falmer Press.

Willms, J. D. (2010). School Composition and Contextual Effects on Student Outcomes. *Teachers College Record*, *112*(4), 1008–1037.

Wilson, D., Burgess, S., & Briggs, A. (2011). The dynamics of school attainment of England's ethnic minorities. *Journal of Population Economics*, *24*(2), 681–700. Retrieved from http://www.jstor.org/stable/41488322 doi: 10.1007/s00148-009-0269-0

Wodtke, G. T., & Parbst, M. (2017). Neighborhoods, Schools, and Academic Achievement: A Formal Mediation Analysis of Contextual Effects on Reading and Mathematics Abilities. *Demography*, *54*(5), 1653–1676. Retrieved from http://link.springer.com/10.1007/s13524-017-0603-1 doi: 10.1007/s13524-017-0603-1

Wolfgramm, C., Morf, C. C., & Hannover, B. (2014). Ethnically based rejection sensitivity and academic achievement: The danger of retracting into one's heritage culture. *European Journal of Social Psychology*, *44*(4), 313–326. doi: 10.1002/ejsp.2018

Xie, Y. (2019). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. Retrieved from https://yihui.name/knitr/

Yang, Y. (2003). Dimensions of Socio-economic Status and their Relationship to Mathematics and Science Achievement at Individual and Collective Levels. *Scandinavian Journal of Educational Research*, *47*(1), 21–41. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00313830308609 doi: 10.1080/00313830308609

Yang, Y., & Gustafsson, J.-E. (2004). Measuring Socioeconomic Status at Individual and Collective Levels. *Educational Research and Evaluation*, *10*(3), 259–288. Retrieved from http://search.proquest.com/docview/62118483?accountid=14026 doi: 10.1076/edre.10.3.259.30268

Yang, Y., Wang, H., Zhang, L., Sean, S., Luo, R., Shi, Y., ... Rozelle, S. (2015). The Han-Minority Achievement Gap, Language, and Returns to Schools in Rural China. *Economic development and cultural change*, *63*(2), 319–359. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25635143

Yang Hansen, K., & Munck, I. (2012). Exploring the measurement profiles of socioeconomic background indicators and their differences in reading achievement: A two-level latent class analysis. *IERI Monograph Series*, *5*, 49–

77. Retrieved from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_05_Chapter_3.pdf

Yang Hansen, K., Rosén, M., & Gustafsson, J.-E. (2011). Changes in the Multilevel Effects of Socioeconomic Status on Reading Achievement in Sweden in 1991 and 2001. *Scandinavian Journal of Educational Research*, *55*(2), 197–211. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00313831.2011.554700 doi: 10.1080/00313831.2011.554700

Yeh, S. (2015). Two Models of Learning and Achievement: An Explanation for the Achievement Gap? *Teachers College Record*, *117*(212), 48.

Zambrano Jurado, J. C. (2013). Análisis multinivel del rendimiento escolar en matemáticas para cuarto grado de Educación Básica Primaria en Colombia. *Revista sociedad y economía*(25), 205–236. Retrieved from http://www.redalyc.org/articulo.oa?id=99629494009

Zhang, L., & Lee, K. A. (2011). Decomposing achievement gaps among OECD countries. *Asia Pacific Education Review*, *12*(3), 463–474. doi: 10.1007/s12564-011-9151-3

Zhang, Z., Parker, R. M. A., Charlton, C. M. J., Leckie, G., & Browne, W. J. (2016). R2MLwiN : A Package to Run MLwiN from within R. *Journal of Statistical Software*, *72*(10), 43. Retrieved from http://www.jstatsoft.org/v72/i10/ doi: 10.18637/jss.v072.i10

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing Multilevel Mediation Using Hierarchical Linear Models. *Organizational Research Methods*, *12*(4), 695–719. Retrieved from http://journals.sagepub.com/doi/10.1177/1094428108327450 doi: 10.1177/1094428108327450

Zhu, L., & Gonzalez, J. (2017). Modeling floor effects in standardized vocabulary test scores in a sample of low SES Hispanic preschool children under the multilevel structural equation modeling framework. *Frontiers in Psychology*, *8*(DEC), 1–14. doi: 10.3389/fpsyg.2017.02146

A | Appendices

A.1 Appendix to Chapter 3

A.1.1 Colombian SES Indices

SES is one of the most important predictors of academic achievement in Colombia, as discussed in subsection 3.4.1. Stratum and Sisben are two indices used to measure SES in Colombia. These have been constructed by the Colombian government to decide who is eligible for public policy programs. Additionally, Colombian National Assessment Institute (Icfes) computes a quality-of-life index called INSE, which is used in their reports. Stratum and Sisben are included in the SABER 11 database, but not INSE. Little is known about how using these indicators affects the conclusions of educational research. This is the question that chapter 7 aims to answer. This subsection reviews the methodology and purpose of these indicators.

A.1.1.1 Stratum

In Colombia, public utilities' rates vary according to the socio-economic status of houses. In order to decide what rate each house should be charged, the government and public utility providers created a socio-economic classification of residential properties called stratum (*estrato*) (Congress of Colombia, 1994b). In the beginning, the way houses were classified in strata was independently decided by each municipality, but since 2004 it is obligatory for all municipalities to use the same methodology and criteria to define strata (Congress of Colombia, 2002), as described below.

In order to decide what stratum each house belongs to, a cluster analysis is carried out for each city in the country, and separately for urban and rural areas within the cities. This analysis aims to find zones in the city in which houses have similar characteristics (i.e. to minimise the within-cluster variation). The variables (house characteristics) that are considered in this analysis are:

- Cadastral information (land use, topography, access roads, land value)
- Type of property (detached or single unit)
- Number of rooms
- Number of bathrooms
- Floor area
- House fittings (bathroom and kitchen characteristics).

The first step in the analysis is to find clusters of zones within each city, using both k-means clustering and a stratification method. Then, the cumulative cluster variance is computed for each of these methods. The method that produces the least cumulative cluster variance is chosen to assign a stratum to each property. Finally, the properties are assigned a stratum according to the zone cluster they belong to. If a property differs from other properties in its cluster (i.e., if the house is an outlier), its stratum is assigned according to its characteristics.

According to the stratum classification, there are six possible strata: properties in stratum 1, 2 and 3 are deprived houses, which can benefit from subsidised utilities' rates; stratum 4 is assigned to middle-class houses, which pay full-price for public services; and strata 5 and 6 are privileged properties, which are charged higher rates in order to subsidise lower strata (DANE, n.d.).

Because of its aim and methodology, the use of stratum as a SES indicator for studying the relationship between socio-economic status and academic achievement may be problematic. Firstly, as mentioned, the aim of classifying properties into strata is to charge different public utilities' rates to advantaged and disadvantaged houses. This implies that a house must benefit from public utilities in order to be assigned to a stratum. In Colombia, 1.2% of the households live in a house without access to any utilities. This figure rises to 5.5% for households in the rural areas (DANE, 2013). Therefore, a non-random part of the population would be excluded from the analysis when using stratum to measure SES.

Secondly, since the cluster analysis for stratum classification is separately performed for each area (urban and rural) and each city in Colombia, identical houses in different cities may be assigned to different strata. This means that the stratum cannot be used to directly compare the SES of students in different areas or cities.

A.1.1.2 Sisben

The aim of the Sisben index is to serve as an instrument for the Colombian government when deciding who is eligible for participation in public policy programs (i.e., for policy focalisation). It consists of a score (ranging from 0 to 100) for each household, which is then discretised to define the groups of potential public policy beneficiaries. Since this is the objective of this index, its score is not computed for all Colombian households, but only for those in potentially vulnerable areas.

The process of assigning a Sisben score to a household consists of two stages. In the first stage, the Colombian Department for National Planning (DNP) selects areas in which information should be collected, which are areas with a large concentration of poor households. In the second stage, households that were not part of the selected survey areas can ask for the DNP to also apply the survey and assign a score (Angulo Sálazar, García Cano, & Velasco Maldonado, 2009). This means that, at the end of 2013, around 70% of Colombian households were assigned a Sisben score, and it is not possible to observe the SES, measured by this indicator, of 30% of the (most advantaged) households in the country (DNP, 2013).

This index is supposed to be updated every three years in order to keep its ability to separate the disadvantaged households from the privileged ones (Congress of Colombia, 2001; DNP, 2008a). Until now, there have been three different versions of the index. The first version of Sisben was implemented in 1995, the second one in 2003 and the last one in September 2011[1] (DNP, 2013; Ministry of Social Protection, 2011). The changes for each version range from the definition of variables and some methodological details, to the reconsideration of the conceptualisation of the index. In the first version, Sisben was thought of as an economic resources index, but it was modified in the second version to become a living-standards index with a multidimensional-poverty approach[2]. The

[1]Here, the implementation date is defined as the moment in which the legislation made the index operable, by stating the thresholds of the index that define eligibility for subsidised health care.

[2]Multidimensional poverty indices are those that consider factors different from income to measure poverty. An example is the Multidimensional Poverty Index (MPI) created by the United Nations Development Programme, which measures poverty using three dimensions: health, education and standard of living (United Nations Development Programme, 2015). In the case of the Sisben III, these dimensions are

statistical methodology has also been changed in each version. For the first version, the index was computed using principal components analysis for four groups of variables (education and social security, income and demographic characteristics, housing quality, and public utilities). In the second version, the principal components analysis was carried out individually for each variable (without aggregating them into groups, as in the first version). The third version of the index adopted a fuzzy sets cluster methodology (DNP, 2013).

Additionally, the thresholds that define who should be public policy beneficiaries have also changed with each version of the index. In the first version, the groups were created with the aim of classifying families as poor and non-poor using different thresholds[3] of unsatisfied basic needs and the extreme-poverty line for urban and rural areas. In the second version, the groups were separately defined for rural and urban zones using k-means clustering. In the last version, each public policy program separately defined the thresholds for potential beneficiaries, instead of setting a single group for all possible policies (DNP, 2013).

Finally, different versions of the Sisben index are computed using different variables, which also change by areas (urban and rural). Table A.1 shows the variables that are used to compute the different versions of the index at each disaggregation level. As can be seen from this table, the third version implied a drastic change by incorporating measures of health and contextual vulnerability, and the second version is the only one that incorporates stratum into the computation of the index. Additionally, in the first version, the indices for both urban and rural areas include the same set of variables, but the thresholds that define who can benefit from social programs is different according to the area. In the second version, both, the thresholds and the index are defined in a different way for each area. The third version applied the fuzzy sets method separately to the group of 14 main cities of the country, as well as the urban and rural areas (DNP, 2013).

Table A.1: Variables included in each Sisben version

Variable	Sisben I (1995)	Sisben II (2003)		Sisben III (2011)		
	National	Urban	Rural	14 cities	Urban	Rural
Health						

Continued on next page

health, education, housing and vulnerability.

[3]For more information on how these thresholds were defined; please refer to DNP and Granados (2003).

Table A.1 – *Continued*

Variable	Sisben I (1995)	Sisben II (2003)		Sisben III (2011)		
	National	Urban	Rural	14 cities	Urban	Rural
Permanent disability				x	x	x
Teenage fertility				x	x	x
Education						
Proportion of functionally illiterate adults				x	x	
Proportion of children not attending school				x	x	x
Being over-aged for school grade		x	x	x	x	x
Proportion of working children				x	x	x
Proportion of adults attaining incomplete secondary or less				x	x	
Attainment for individuals aged 11 or older	x					
Attainment for the person with the highest income	x					
Attainment for the head of the household		x	x			
Attainment of the head of household's spouse		x	x			
Housing						
Dwelling type (house or flat, room, other, or Indigenous house)				x	x	x
Water source	x	xs	xs		x	
Toilet connection type	x	xs	xs	x		
Shared toilet				x	x	
Number of toilets		xs	xs			
Toilet location		xs	xs			
Shower			x			
Flooring material	x	xs	xs	x		
Wall material	x		x	x		
Garbage collection	x	xs	xs			
Energy source for cooking		xs	xs	x		
Overcrowding	x	x	x			
Roof material	x					
Stratum		x	x			
Vulnerability						
Individual						

Continued on next page

Table A.1 – *Continued*

Variable	Sisben I (1995)	Sisben II (2003)		Sisben III (2011)		
	National	Urban	Rural	14 cities	Urban	Rural
Household size						x
Head of household				x	x	x
Age dependency ratio				x	x	x
Assets*	x**	x	x	x		
Landline		xs	xs			
Exclusive use of landline		xs	xs			
Proportion of children under 6	x					
Proportion of working individuals	x	xs				
Income per capita	x					
Social security of the person with the highest income	x					
Proportion of household members in the contributory health regime		x	x			
Contextual (at the municipality level)						
Child mortality rate				x	x	x
Homicide rate				x	x	x
Net school enrolment ratio by school level				x	x	x
Use of health services				x	x	x

x: The variable is included in this version of the index.

xs: The variable is included as an interaction with stratum.

*: Assets include fridge, washing machine, T.V., cable, water heater, microwave oven, air conditioner, computer, sound system, motorbike, tractor, car and real estate.

**: Number of assets.

Note: Data from DANE (2013).

When the objective is to understand how SES interacts with academic achievement for all students in the country, the changing nature of Sisben and its focus on the most deprived households introduce comparability and self-selection problems. These keep Sisben away from being an ideal indicator for comparing SES of students in different regions and cohorts.

A.1.1.3 INSE

Since 2007, Icfes computes the INSE index, a quality-of-life indicator that is included into reports at the national and LA levels. This index is the resulting principal component of PCA after transforming categorical variables into continuous variables. The specific variables that are included in the analysis change each term (twice a year), but are grouped into four categories: housing, human capital, utilities and overcrowding. Besides parental education and occupation and family income, INSE includes the stratum and Sisben indices and a set of durable items (Icfes, 2010a).

Although this indicator is not included as a variable in the SABER 11 database, it could be computed using the available variables. Nonetheless, as it is the case for the stratum and Sisben indicators, it is unknown how using this indicator would affect the conclusions about the ethnic achievement gap in Colombia. That is why chapter 7 evaluates competing index-creation techniques to evaluate their effect in the context of this thesis.

A.2 Appendix to Chapter 4

A.2.1 Data Cleaning Procedure

The database used in this dissertation results from merging two databases:

1. Colombian Ministry of Education (MoE)'s 2016 Buscando Colegio (BC) (Ministry of Education, 2016a), which includes both, an institution identifier from the National Bureau of Statistics (DANE, its acronym in Spanish) for 32,019 institutions, and a DANE's school-venue identifier for 62,782 school venues. This venue-level dataset also includes information on location, grades offered, teaching model, and status for each school venue. The dataset includes closed schools, and schools offering all education levels (preschool, elementary, secondary and middle school).

2. Icfes' SABER 11 2008-2013, which provides demographic and socio-economic information, as well as test scores for 6,661,170 students who took the SABER 11 test between 2008 and 2013 and two school identifiers:

   (a) An identifier from DANE (DANE ID), which may be related either to an institution or to a school venue, but this is not clear from the dataset manuals. There are 10,653 unique identifiers of this kind in this dataset.

   (b) An Icfes identifier, which classifies students as belonging to different groups within the same school venue, if they have a different school calendar, school day or school focus[4]. This dataset identifies 13,684 school groups that the students belong to.

   Table A.2 shows how these identifiers are distributed over time, as well as the number of students that were originally included in the database.

As shown in this table, the number of students and schools have been steadily increasing. However, there is a sharp drop of students in 2013, when a nation-wide strike in many sectors caused logistic complications in many parts of the country (El Espectador, 2013).

[4]As the described below, the analysis focuses on the institution level. When there were different school days or focus within the institution, the one that was taken by more students was picked as the institution-level characteristic.

Table A.2: Number of students and unique DANE and Icfes identifiers per year

	2008	2009	2010	2011	2012	2013	Total
Students	506,427	524,809	570,765	572,197	580,564	575,823	3,330,585
Icfes ID	10,160	10,554	11,102	11,725	12,025	12,177	13,684
Venues ID	8,535	8,818	9,127	9,535	9,780	9,925	10,603
Institutions ID	8,429	8,698	8,987	9,373	9,605	9,740	10,653
LAs	80	84	94	94	94	94	94

School groups in the SABER 11 database were matched with school venues and institutions in the BC database in six steps, in which the identifiers used for matching were changed. Table A.3 shows the BC database's identifiers that were used, and the number of students, school groups, school venues and institutions that were identified on each step for all the years in our database.

The SABER 11's DANE ID were matched with both, the BC database DANE's school-venue identifier (first step) and the BC database DANE's institution identifier (third step). The SABER 11 database's school name and city were also matched[5] with the BC's venue's name and city (second step) and with the BC's institution's name and city (fourth step). Finally, the SABER 11's school name was matched with the BC's institution (fifth step) and venue name (sixth step).

Table A.3: Merge process

Identifier	2008	2009	2010	2011	2012	2013
Students	498,440	517,411	563,116	566,025	575,590	572,830
Venue's ID	492,885	511,443	556,210	559,749	569,109	566,491
Venue's name and city	3,843	4,300	4,816	4,615	4,918	5,120
Institution's ID	55	85	110	94	95	83
Institution's name and city	805	659	927	869	689	603
Institution's name	526	575	755	592	586	407
Venue's name	326	349	298	106	193	126
Groups	9,963	10,376	10,936	11,564	11,902	12,107
Venue's ID	9,818	10,229	10,771	11,400	11,731	11,944
Venue's name and city	101	104	117	113	126	125
Institution's ID	3	3	4	4	4	4
Institution's name and city	21	19	24	27	21	16

Continued on next page

[5] as it is not possible to know in advance whether they belong to the school venue or to the institution

Table A.3 – *Continued*

Identifier	2008	2009	2010	2011	2012	2013
Institution's name	13	15	15	16	13	12
Venue's name	7	6	5	4	7	6
Venues	8,534	8,817	9,126	9,534	9,779	9,924
Venue's ID	8,410	8,690	8,984	9,388	9,630	9,788
Venue's name and city	83	87	98	96	108	99
Institution's ID	3	3	3	3	4	4
Institution's name and city	21	19	24	27	21	16
Institution's name	13	14	13	16	11	12
Venue's name	4	4	4	4	5	5
Institutions	8,428	8,697	8,986	9,372	9,604	9,739
Venue's ID	8,307	8,573	8,849	9,232	9,459	9,608
Venue's name and city	82	87	96	95	106	97
Institution's ID	2	2	2	2	3	3
Institution's name and city	20	17	22	23	20	14
Institution's name	13	14	13	16	11	12
Venue's name	4	4	4	4	5	5

As can be seen from the table, the proportion of students, groups, venues and institutions that are matched using each identifier is relatively constant over time. In particular, around 98.9% of the observations are matched using the DANE's venue identifier each year.

After this process, the database identifies a set of cross-subsections with a total of 3,293,412 students in 13,394 school groups, 10,602 school venues, 10,590 institutions and all LA. This implies that it was not possible to identify the school venue and institution for 290 school groups, affecting 37,173 students in the sample. Table A.4 shows how these are distributed over time. As shown in the table, the proportion of unmatched observations is higher for earlier years. For example, while in 2008 1.6% of students cannot be matched to the BC database, this proportion decreases to 0.5% in 2013. Overall, 1.1% of the students and 2.1% of the groups in the original SABER 11 database cannot be matched.

A closer look to the SABER 11 data shows that 1.5% of the students attend institutions that run under programs for students with distinct scientific, artistic or sport abilities. Also 0.6% of the students attend institutions offering education for disabled students under a non-traditional program. Additionally, 9.5% of the students are part of

Table A.4: Proportion (%) of unmatched students, groups, venues and institutions per year

	2008	2009	2010	2011	2012	2013	Total
Students	1.58	1.41	1.34	1.08	0.86	0.52	1.12
Groups	1.94	1.69	1.50	1.37	1.02	0.57	2.12
Venues	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Institutions	0.01	0.01	0.01	0.01	0.01	0.01	0.59
LAs	0.00	0.00	0.00	0.00	0.00	0.00	0.00

school groups offering instruction at evening or weekend. Finally, 0.1% of the students attend one institutions with less than 5 students taking the exam on the respective year. These students are excluded from the sample, since they are probably not representative of the majority of Colombian students.

Besides, 9.4% of the students have incomplete data (missing values for any variable in the analysis) and they are also excluded from analysis. Table A.5 shows how the proportion of excluded observations is distributed over time and education-system level. In particular, it shows that the exclusion criteria affects all levels evenly over time, although a more observations are dropped in 2008 when considering the variables used for the analysis.

Table A.5: Proportion (%) of deleted observations by reason and year

Level	2008	2009	2010	2011	2012	2013	Total
Students							
Talent Program	1.62	1.58	1.48	1.55	1.56	1.52	1.55
Disability Program	0.48	0.53	0.60	0.61	0.68	0.64	0.59
Alternative School Day	7.76	8.69	9.83	10.29	10.22	10.17	9.54
Small school size	0.11	0.09	0.08	0.08	0.08	0.09	0.09
Incomplete Data	14.40	8.06	8.04	8.06	11.44	6.75	9.38
Groups							
Talent Program	1.42	1.39	1.35	1.34	1.30	1.28	1.34
Disability Program	0.59	0.66	0.63	0.71	0.72	0.78	0.69
Alternative School Day	7.92	8.56	9.24	9.81	9.35	9.87	9.17
Small school size	1.76	1.52	1.41	1.31	1.33	1.39	1.44
Incomplete Data	13.89	8.12	8.09	7.47	10.51	6.24	8.95
Venues							

Continued on next page

Table A.5 – *Continued*

Level	2008	2009	2010	2011	2012	2013	Total
Talent Program	1.37	1.35	1.35	1.29	1.26	1.24	1.31
Disability Program	0.68	0.76	0.76	0.84	0.86	0.92	0.81
Alternative School Day	7.37	7.80	8.39	8.70	8.36	8.45	8.20
Small school size	2.04	1.79	1.69	1.58	1.62	1.69	1.73
Incomplete Data	13.59	8.25	8.20	7.22	10.54	6.32	8.94
Institutions							
Talent Program	1.34	1.32	1.32	1.27	1.24	1.22	1.28
Disability Program	0.63	0.73	0.75	0.80	0.83	0.88	0.78
Alternative School Day	7.38	7.80	8.37	8.66	8.30	8.39	8.17
Small school size	2.06	1.82	1.71	1.61	1.65	1.73	1.76
Incomplete Data	13.59	8.32	8.27	7.20	10.56	6.35	8.97

There are several patterns that become evident from this table. First, the proportion of institutions offering especial programs for talented students (those with with distinct scientific, artistic or sport abilities) has been decreasing over time; second, the proportion of institutions offering special programs for disabled students has been increasing, as well as the proportion of students attending schools in alternative school days (at evening or weekend); and the largest proportion of the deleted observations responds to alternative school days. After deleting the observations according to these reasons, we lose 20.4% of the students, 14.2% of the groups, 12.5% of the venues and 14.8% of the institutions in the matched dataset. Table A.6 shows the distribution of deleted observations by year and level. Reflecting the patterns we observe, the proportion of missing observations increases over time

Table A.6: Proportion (%) of dropped students, groups, venues and institutions per year

	2008	2009	2010	2011	2012	2013	Total
Students	23.43	18.53	19.39	19.79	22.82	18.34	20.35
Groups	16.69	15.65	15.74	15.93	15.28	15.25	14.16
Venues	15.31	14.11	14.33	14.33	14.11	14.06	12.49
Institutions	15.26	14.04	14.25	14.23	13.99	13.92	14.77
LAs	0.00	0.00	0.00	0.00	0.00	0.00	0.00

A.2.2 Data Structure

Figure A.1 shows the structure of the data by plotting an histogram for the number of school venues within institutions (first row), and another histogram for the number of groups within school venues (second row) for each year in our database. As shown in this figure, every year both distributions are skewed to the right, which means that most institutions consist of a unique group within a school venue. Specifically, around 98.5% of institutions only have one school venue, and 83.2% of school venues only have one group. Furthermore, the maximum number of school venues within an institution is 6, the same as the maximum number of groups within a school venue. , both maximums reached in 2012.

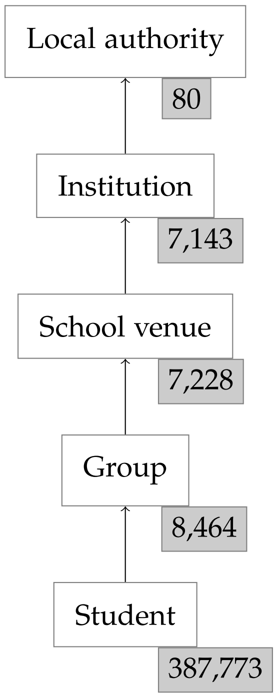

Figure A.1: Data structure

In order to further understand our data structure, the first column of Figure A.2 shows a boxplot of the size of the groups, school venues and institutions, in terms of the number of students within each level. Boxplots represent the distribution of a variable through vertical lines, which correspond to (from left to right):

- the lower adjacent value, which is the smallest value of the variable that is greater than or equal to $Q_1 - 1.5\,(Q_3 - Q_1)$

- the first quartile or 25th percentile, Q_1

- the median, second quartile or 50th percentile, Q_2

- the third quartile or 75th percentile, Q_3

- the upper adjacent value, which is the largest value of the variable that is smaller than or equal to $Q_1 + 1.5\,(Q_3 - Q_1)$.

The points succeeding the last (or preceding the first) vertical lines represent outliers. These are observations that are too far away from the interquartile range and, therefore, far from most data points.

The second column of Figure A.2 zooms in on the distributions of the size of the groups, school venues and institutions, by showing the histograms of these variables up to their upper adjacent value. As shown in this figure, these distributions are similar for all cohorts, as all of them are skewed to the right. Groups, school venues and institutions have 50.6, 60.9 and 56.4, students on average, respectively. However, these averages are influenced by the long tails of the distributions. The median size of groups is 36 students, 38 for school venues and 36 for institutions.

This analysis leads to the cpnclusion that the division between school groups, venues and institutions is redundant since, in general, students are nested within a single group, which is also nested within a single school venue and an institution. For this reason, institutions are considered the sole higher level classification and are referred as schools.

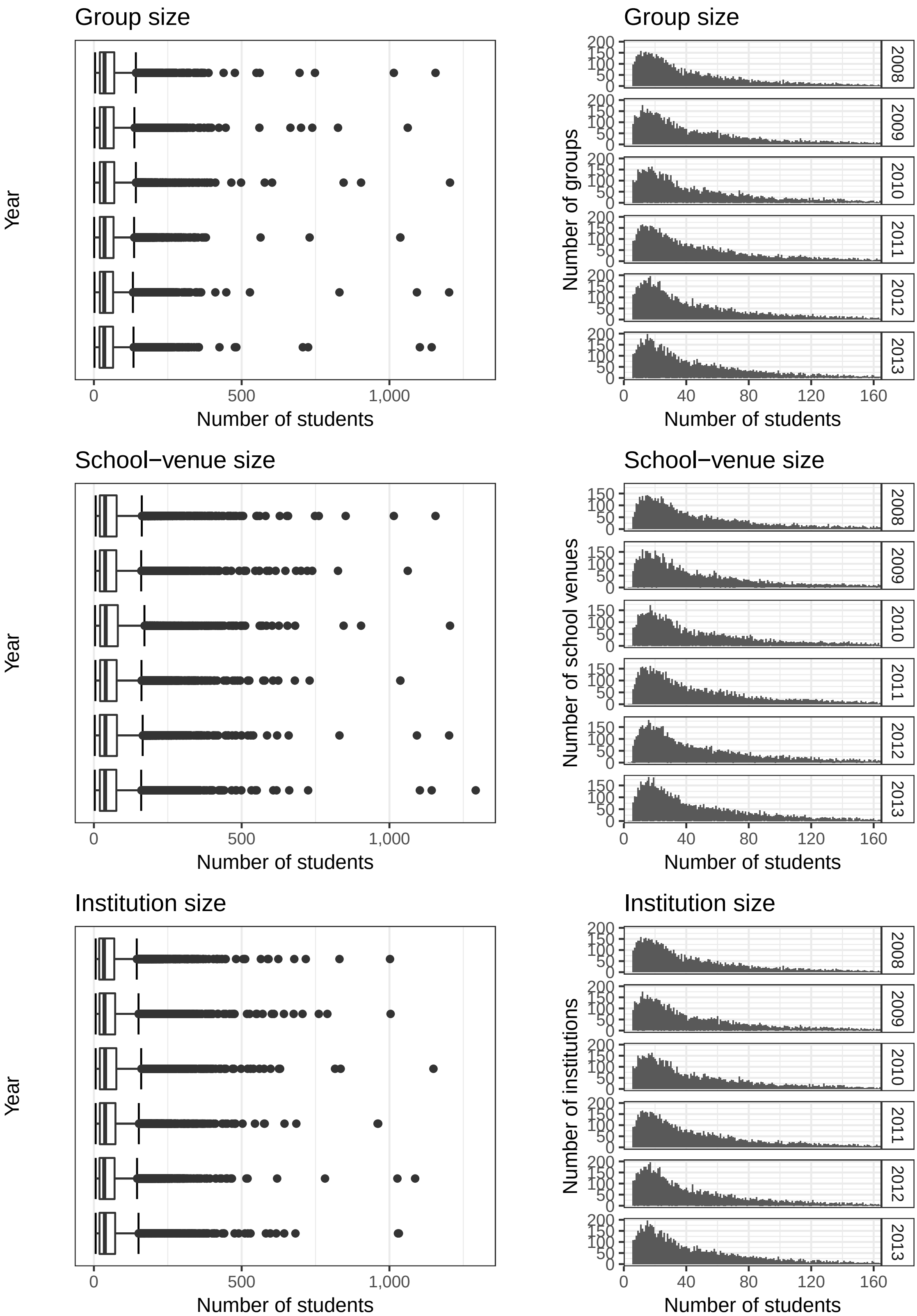

Figure A.2: Distributions of the size of institutions, school venues and groups

A.2.3 Additional Information about the SABER 11 Exam

This subsection examines the raw SABER 11 maths test scores over time. This allows to understand the anomalies in them that affect their comparability over time. The SABER 11 test scores are scaled to have mean 50 and standard deviation 10 in 2000, when the competency-based version of the exam was created (Icfes, 2013). There are also two versions of the exam per year, one around April (denoted by number 1) and a second one around September (denoted by number 2). Each of these applications have a different set of questions that are, in theory, comparable.

Figure A.3 shows that the test scores behave differently before and after 2010, and both applications in 2010 are different to any other applications. Between 2008 and 2009, around 50,600 students per year took the exam in April and 357,000 students per year took the exam in September. Besides 51% of the around 1,050 schools in each of April's applications were private, while 31% of the around 6,250 schools in each of September's applications were so. That is, most students and schools (especially state schools) in the country used to take the exam in the second semester of the year, as they run under school year type A. Students in certain regions, where the school-year type B and private schools dominate, used to take the exam in the first semester of the year (in April).

In contrast, between 2011 and 2013, after all state schools changed to school-year type A, around 18,400 students per year took the exam in April and 440,500 students per year took the exam in September. Besides, all of the around 600 schools in each of April's applications were private, while 20% of the around 11,700 schools in each of September's applications were so. Therefore, differences in the composition of test-takers in both applications during the year may explain the change in the mean test scores gap between applications in the same year. There is a difference of 1.3 points between the applications in 2008, while the difference between the applications in 2013 is of 8.5 points. Besides, the mean scores for both applications are higher after 2010.

Finally, 2010 is a special year for two reasons: first, it is the first year in which all state schools take the exam in September, which implies that it is the year in which the population of test takers drastically decreased during the first application of the exam, and increased during the second application of the year. Second, a change in the flexible component of the exam was introduced during the first application of the year, reducing in 15 the total number of questions of the exam (Icfes, 2009, 2010b). In

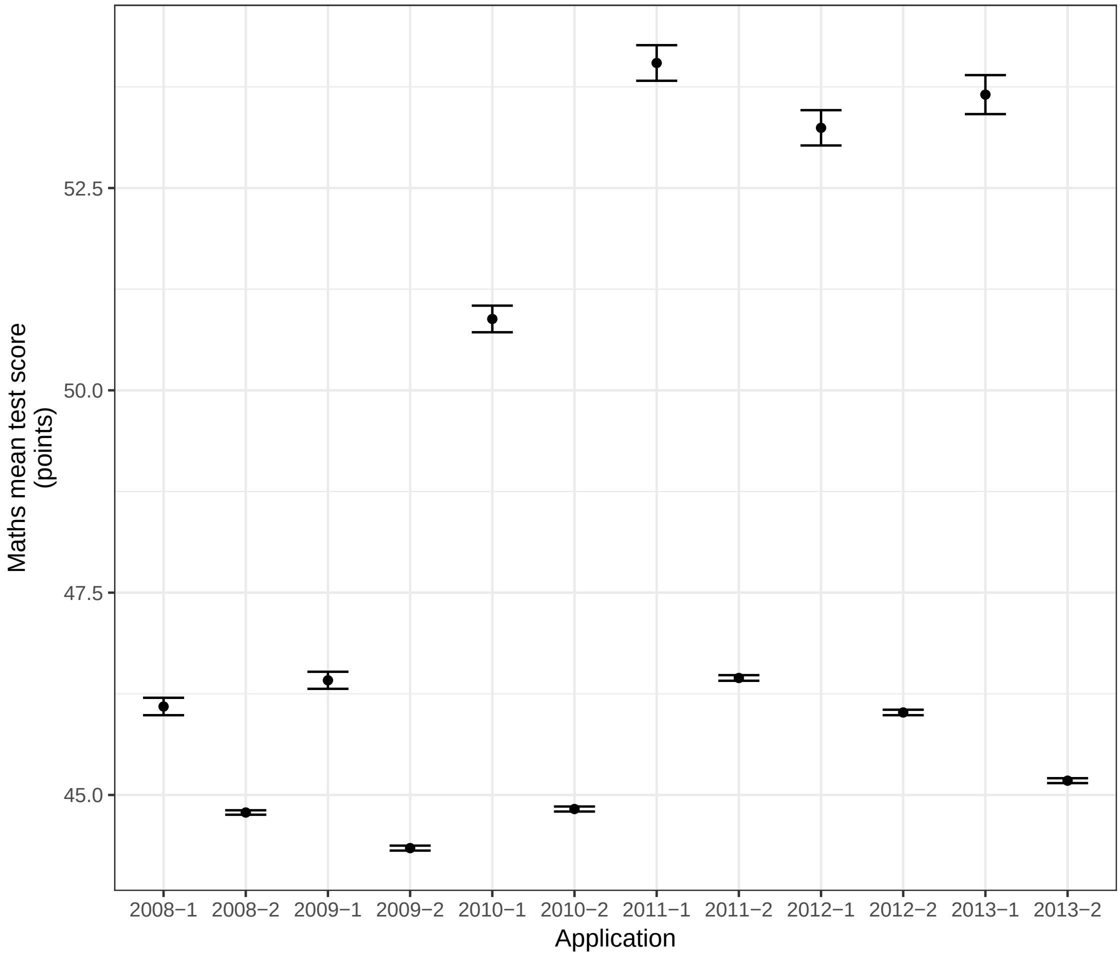

Figure A.3: Mean maths test scores by application

theory, this change should not have affected the test scores of the core subjects of the exam (including the maths test scores).

The population of test-takers in each application of the year is similar to the one after 2010, but the gap in the mean test score between applications in 2010 is closer to the gap between 2008 and 2009. Similarly to the applications after 2010, 18,160 students took the exam in April and 441,947 students took the exam in September that year; 98.2% of the 444 schools in April's application were private, while 29.7% of the 7,262 schools in September's application were so. However, the gap in the mean maths test score between applications is 6.1 points. This implies that differences in the test-takers composition is not the only explanation for the changes in the test scores before and after 2010.

Given this anomaly, it is evident that SABER 11 test scores cannot be directly compared over time. One option would be to normalise each exam application, but given the differences in the school-type composition for the applications in the same year, it could distort the magnitude of the mean differences between private and state schools. Besides, in order to examine differences between state and private schools, it is necessary to simultaneously consider the two applications every year. Therefore, the test scores are normalised per year, and the analysis is restricted to years in which both applications within the year have the same characteristics, so they can be treated as a single unit.

A.2.4 Supplementary Tables and Figures

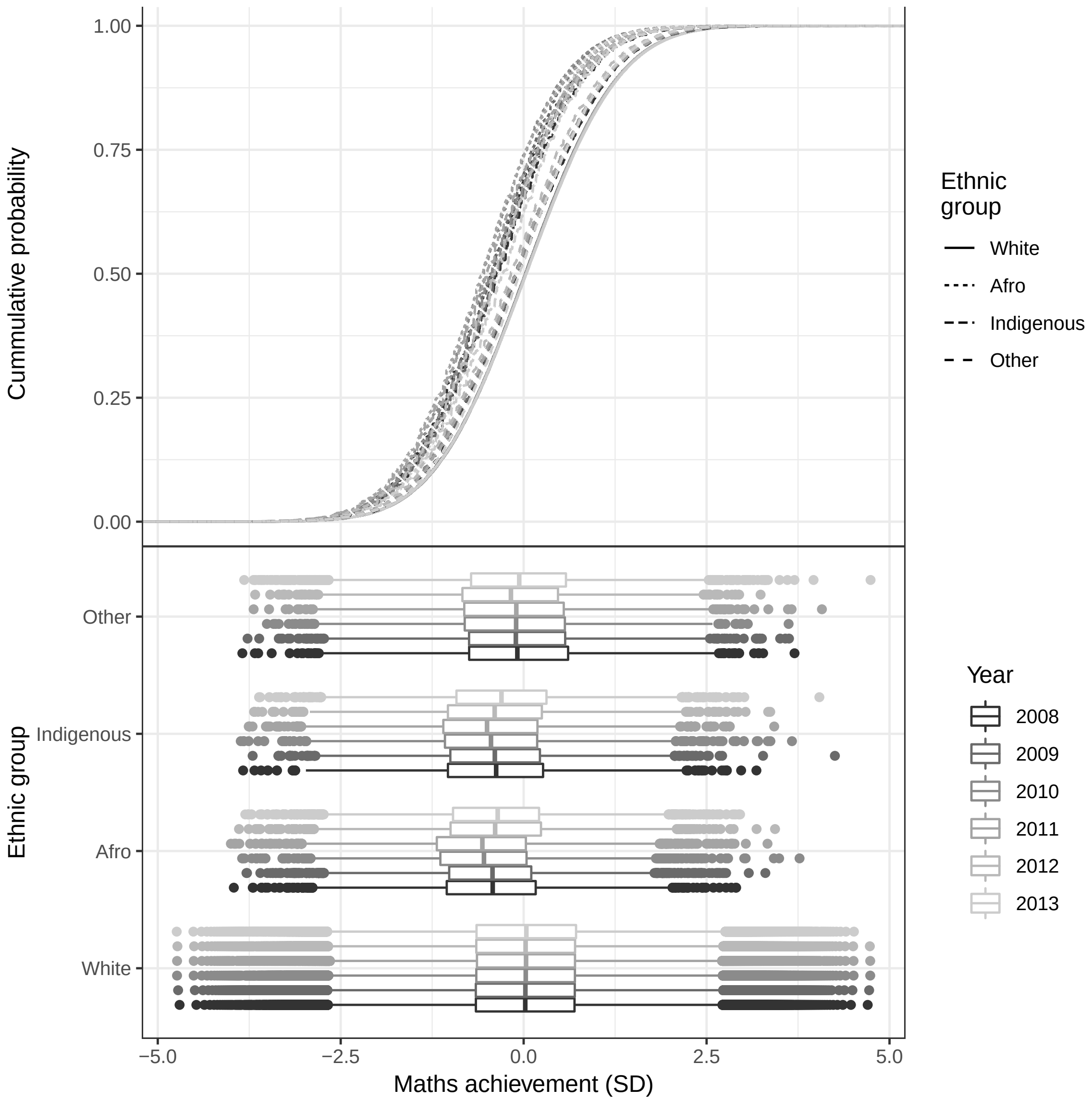

Figure A.4: Cumulative distribution of maths test scores over time and by ethnic group

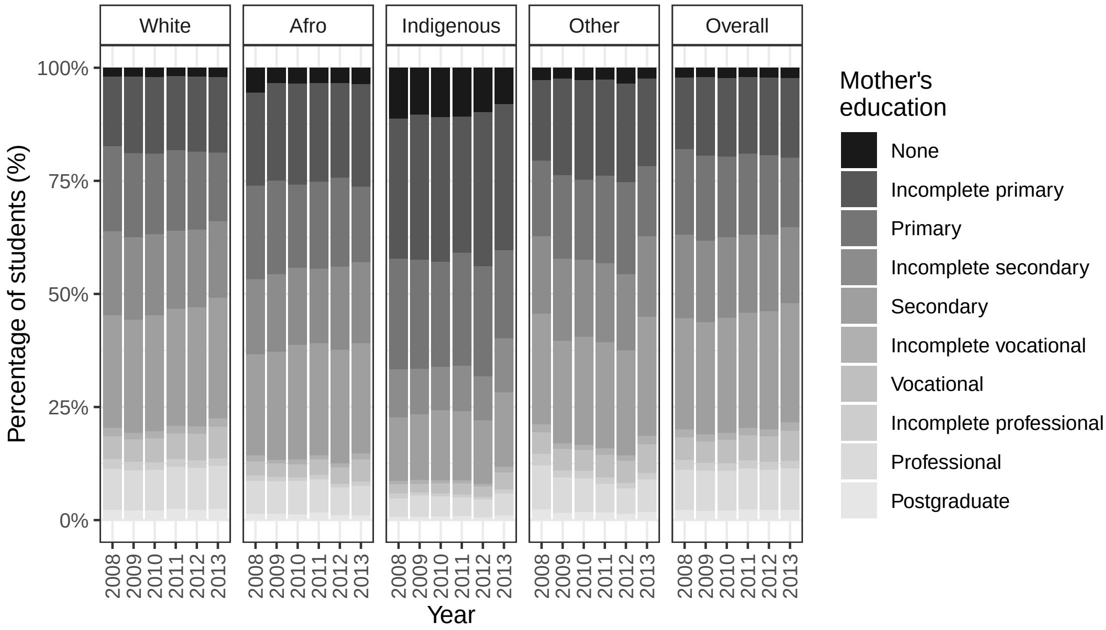

Figure A.5: Distribution of the mother's educational attainment by ethnicity over time

Table A.7: Percentage (%) of girls in the sample by ethnic group and year

Year	Overall	White	Afro	Indigenous	Other
2008	54.54	54.65	54.71	49.28	52.77
2009	54.58	54.67	55.27	49.93	51.58
2010	54.63	54.73	55.35	50.08	51.61
2011	55.13	55.11	57.90	52.75	53.10
2012	54.90	54.89	56.67	51.73	55.23
2013	54.61	54.58	56.68	52.47	54.32

Table A.9: Mean school proportion of students of each ethnic group (%) according to the school characteristics

Variable	Group	2008	2009	2010	2011	2012	2013
Type							
State	White	89.49	89.77	89.35	88.90	88.08	81.78
Private	White	96.40	96.96	96.81	96.75	97.02	91.85
State	Afro	5.04	5.26	5.23	5.38	5.99	7.40
Private	Afro	1.43	1.27	1.51	1.57	1.62	2.85
State	Indigenous	3.17	3.24	3.25	3.46	3.90	4.77

Continued on next page

Table A.9 – *Continued*

Variable	Group	2008	2009	2010	2011	2012	2013
Private	Indigenous	0.35	0.41	0.48	0.40	0.42	0.81
State	Other	2.31	1.73	2.16	2.26	2.03	6.05
Private	Other	1.82	1.36	1.20	1.29	0.93	4.50
Zone							
Urban	White	94.65	95.16	94.80	94.79	94.80	89.15
Rural	White	84.68	85.03	84.99	84.86	83.25	76.99
Both	White	91.11	91.50	91.15	90.37	91.08	85.37
Urban	Afro	2.48	2.48	2.69	2.70	2.96	4.26
Rural	Afro	6.78	6.57	6.43	6.36	7.26	8.82
Both	Afro	4.85	5.69	5.36	5.83	5.71	6.57
Urban	Indigenous	0.91	0.91	0.98	0.97	0.95	1.47
Rural	Indigenous	5.77	6.12	5.89	5.90	6.85	7.75
Both	Indigenous	2.16	1.65	1.81	2.02	1.97	2.76
Urban	Other	1.96	1.45	1.53	1.55	1.29	5.12
Rural	Other	2.77	2.27	2.69	2.87	2.64	6.43
Both	Other	1.88	1.16	1.68	1.78	1.24	5.29
Day type							
Full day	White	95.78	96.08	95.60	95.81	95.98	91.63
Morning	White	89.85	90.22	89.94	89.52	88.56	81.97
Afternoon	White	91.64	91.82	91.65	90.70	91.19	84.25
Full day	Afro	1.07	1.12	1.09	1.10	1.08	1.48
Morning	Afro	5.08	5.20	5.27	5.37	6.20	7.96
Afternoon	Afro	4.56	4.83	5.08	5.56	5.16	7.18
Full day	Indigenous	1.47	1.63	1.91	1.63	1.95	2.31
Morning	Indigenous	2.70	2.70	2.65	2.93	3.20	4.15
Afternoon	Indigenous	1.61	1.88	1.77	1.89	2.15	2.50
Full day	Other	1.68	1.17	1.40	1.47	0.99	4.58
Morning	Other	2.37	1.87	2.14	2.19	2.04	5.92
Afternoon	Other	2.18	1.47	1.50	1.84	1.50	6.07
Focus							
Academic	White	93.35	93.73	93.28	93.16	92.52	87.01
Academic and Technical	White	92.09	92.99	92.65	92.70	93.00	85.80
Teaching	White	88.14	87.07	91.67	88.09	87.95	81.65

Continued on next page

Table A.9 – *Continued*

Variable	Group	2008	2009	2010	2011	2012	2013
Technical	White	88.07	88.08	87.05	86.39	85.95	79.43
Academic	Afro	2.88	3.07	3.19	3.14	3.59	4.85
Academic and Technical	Afro	4.28	3.77	4.12	4.40	4.55	6.17
Teaching	Afro	5.73	6.14	5.15	7.11	6.87	6.91
Technical	Afro	5.80	6.17	6.12	6.49	7.14	8.75
Academic	Indigenous	1.70	1.78	1.89	2.03	2.38	3.00
Academic and Technical	Indigenous	1.35	1.42	1.34	1.13	1.19	2.39
Teaching	Indigenous	4.27	4.34	1.31	2.15	2.04	3.74
Technical	Indigenous	3.79	3.95	4.40	4.34	4.66	5.31
Academic	Other	2.07	1.42	1.64	1.67	1.51	5.14
Academic and Technical	Other	2.28	1.81	1.89	1.76	1.26	5.64
Teaching	Other	1.87	2.46	1.88	2.66	3.14	7.71
Technical	Other	2.34	1.80	2.43	2.77	2.25	6.51
Ethnoeducation							
Yes	White	21.94	25.43	29.25	26.43	24.81	16.51
No	White	92.87	93.19	92.93	92.70	92.46	86.60
Yes	Afro	38.85	34.26	28.78	29.44	28.25	30.94
No	Afro	3.29	3.46	3.55	3.64	4.02	5.34
Yes	Indigenous	26.71	28.58	30.56	32.63	36.20	39.18
No	Indigenous	1.85	1.89	1.84	1.90	2.04	2.70
Yes	Other	12.49	11.73	11.42	11.51	10.74	13.37
No	Other	1.99	1.46	1.68	1.77	1.47	5.36

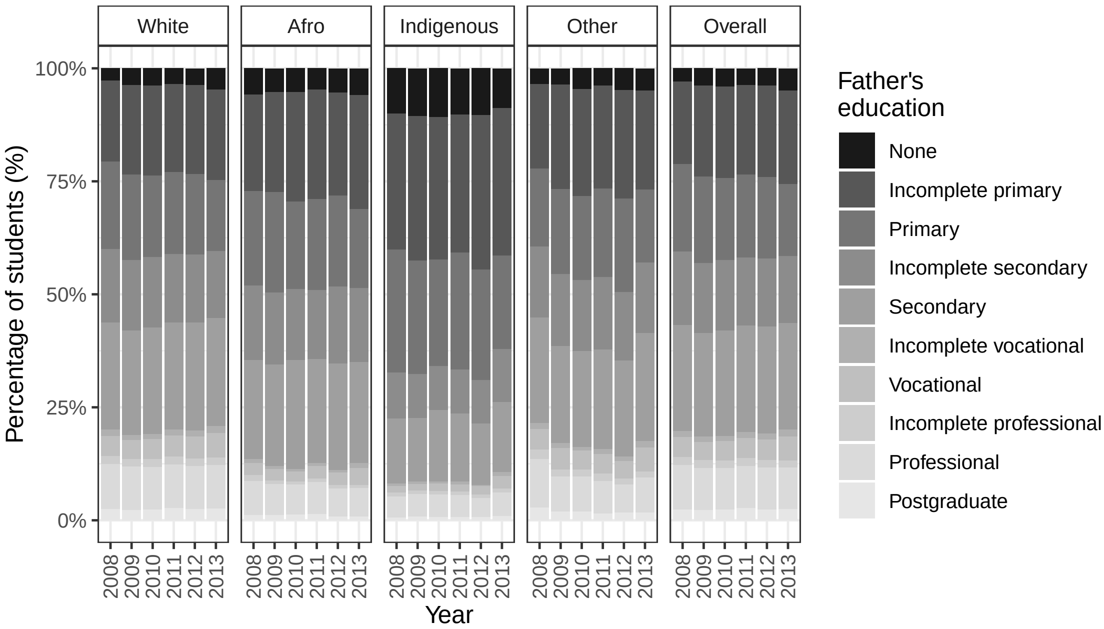

Figure A.6: Distribution of the father's educational attainment by ethnicity over time

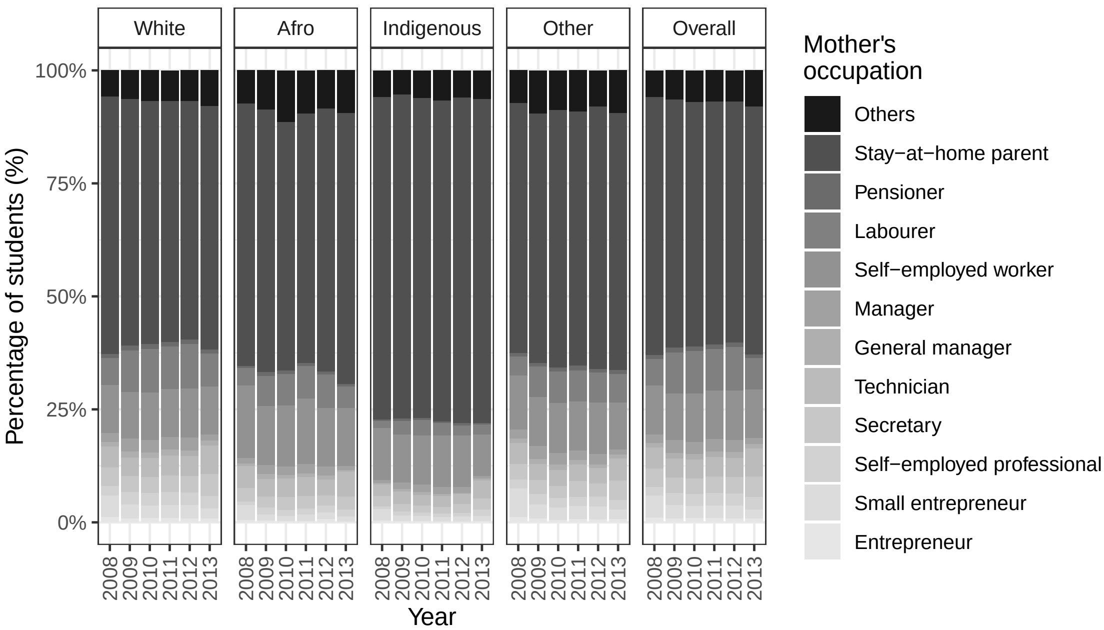

Figure A.7: Distribution of the mother's occupational status by ethnicity over time

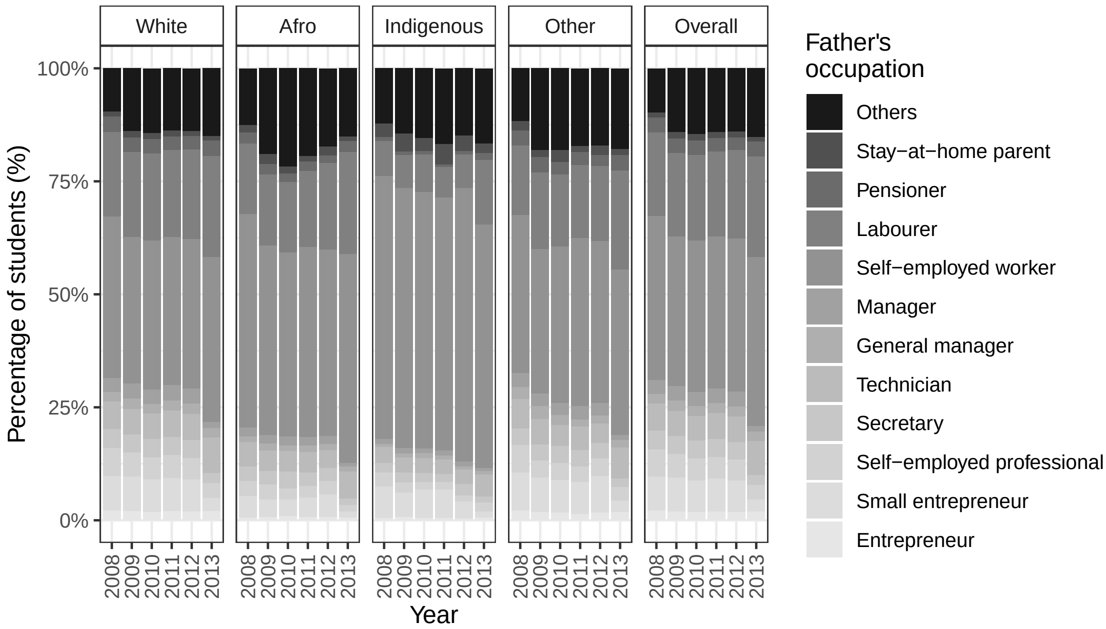

Figure A.8: Distribution of the father's occupational status by ethnicity over time

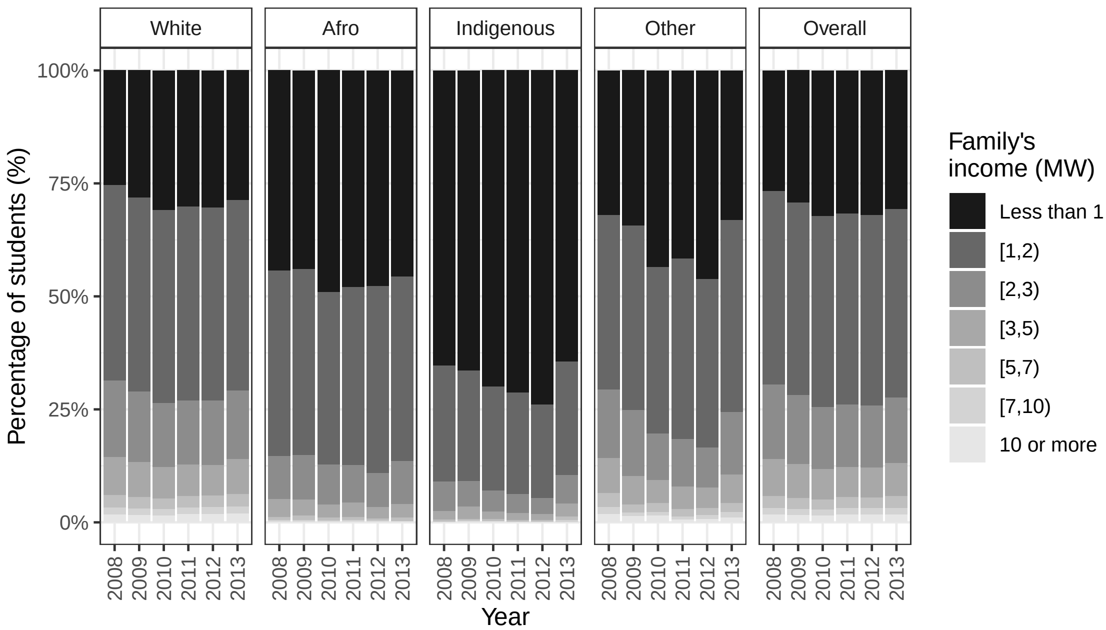

Figure A.9: Distribution of the family's income by ethnicity over time

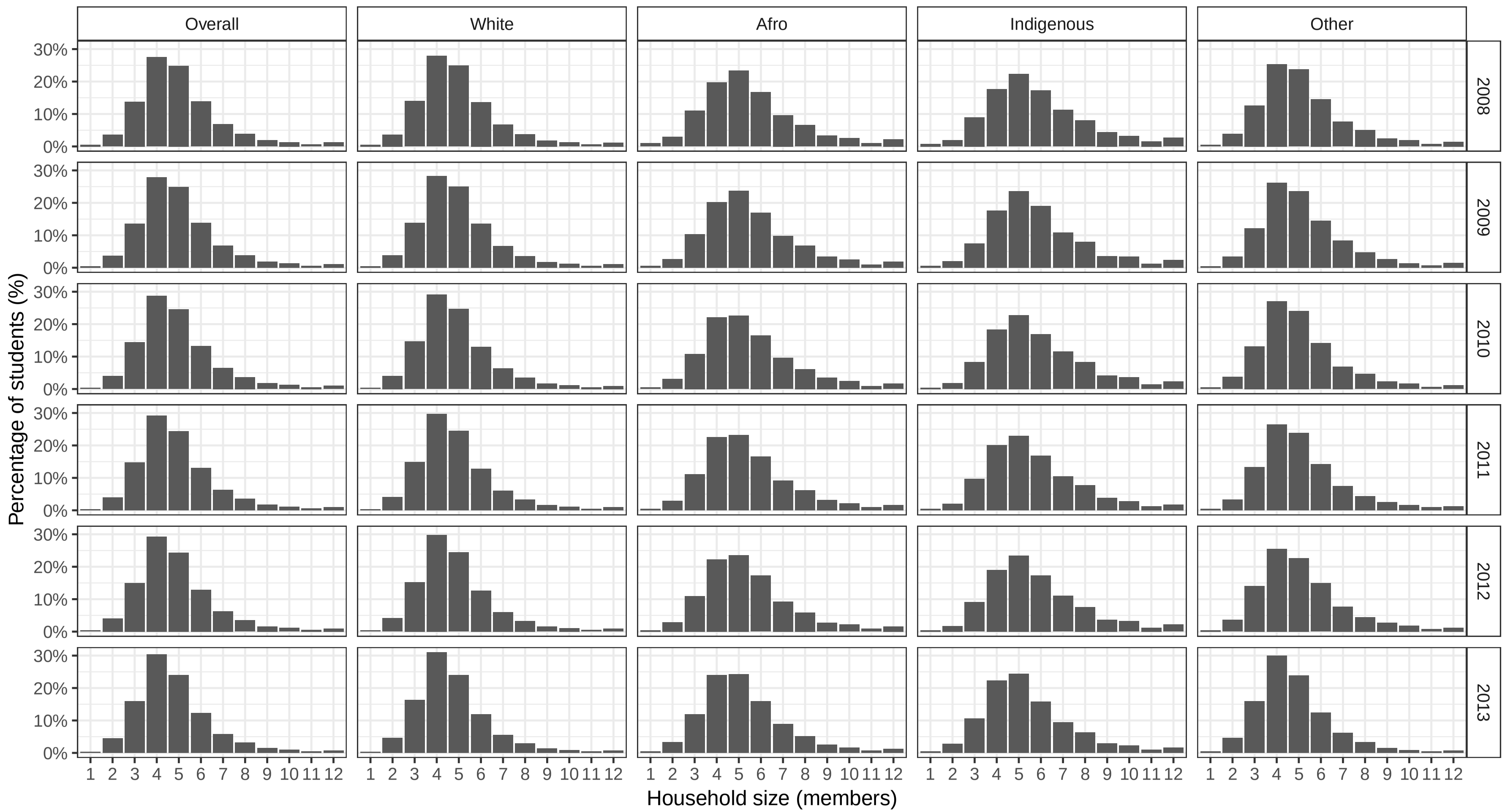

Figure A.10: Distribution of the student's household size by ethnicity over time

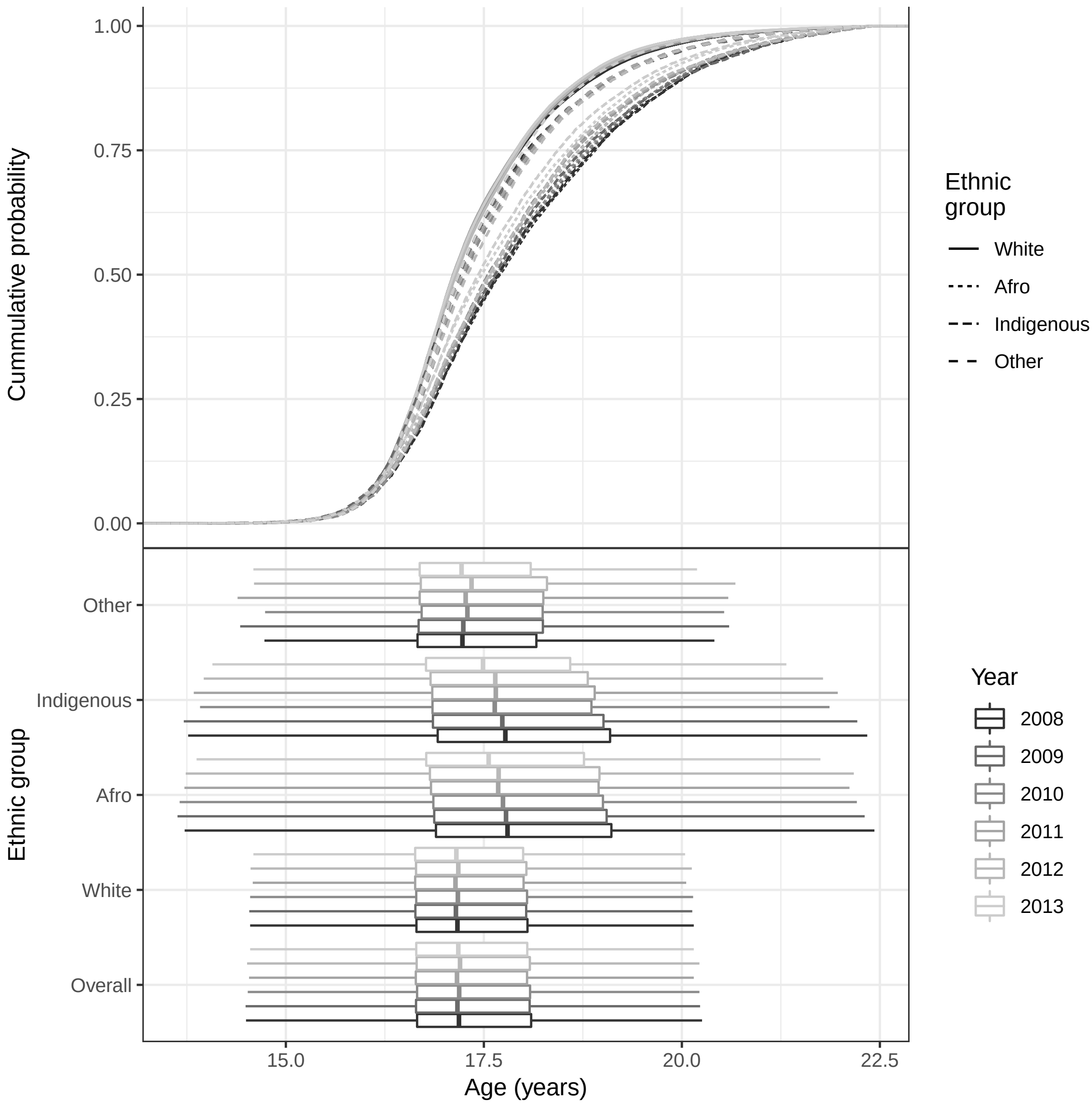

Figure A.11: Distribution of the student's age by ethnicity over time

Table A.8: Frequency of each school participation pattern in the final dataset

Pattern	Frequency	%	Cummulative %
111111	6,199	68.27	68.27
.11111	423	4.66	72.93
...111	288	3.17	76.10
..1111	283	3.12	79.22
.....1	276	3.04	82.26
....11	246	2.71	84.97
11111.	179	1.97	86.94
1.1111	112	1.23	88.17
11.111	89	0.98	89.15
1111.1	78	0.86	90.01
111.11	64	0.70	90.72
1111..	58	0.64	91.35
1.....	55	0.61	91.96
....1.	53	0.58	92.54
.1.111	51	0.56	93.11
Other patterns	626	6.89	100.00

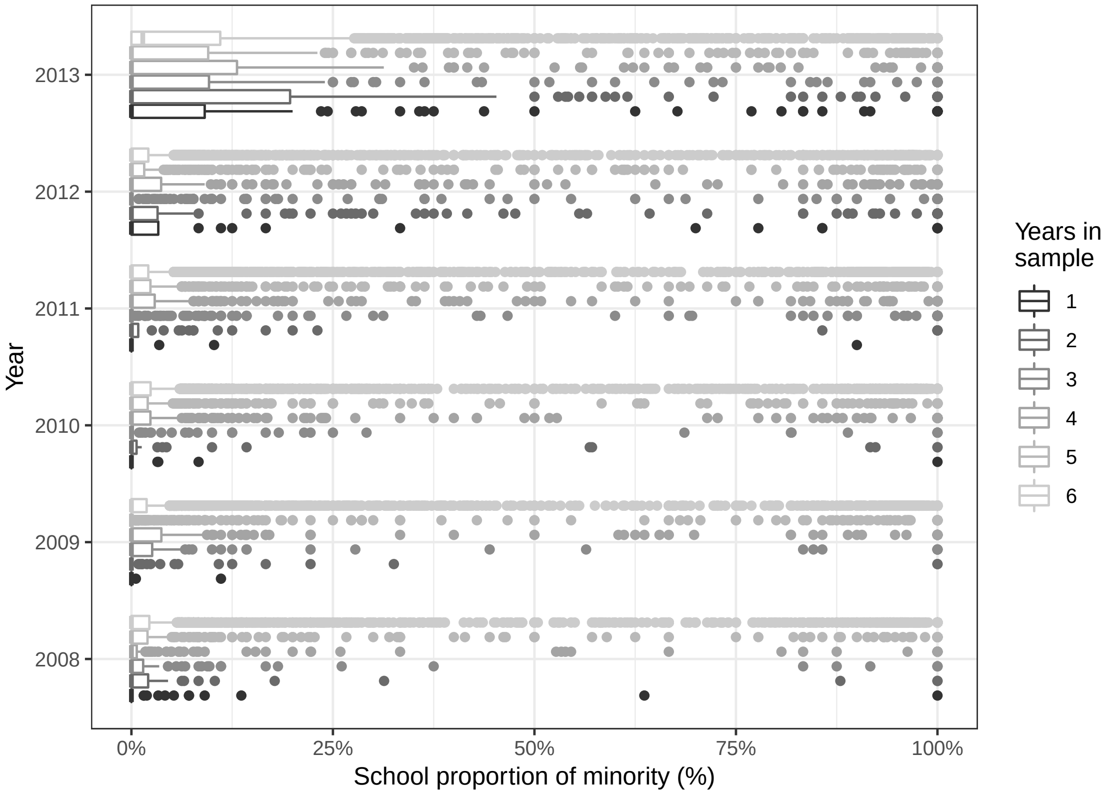

Figure A.12: Distribution of the schools' proportion of minority students according to the number of years they appear on the sample over time

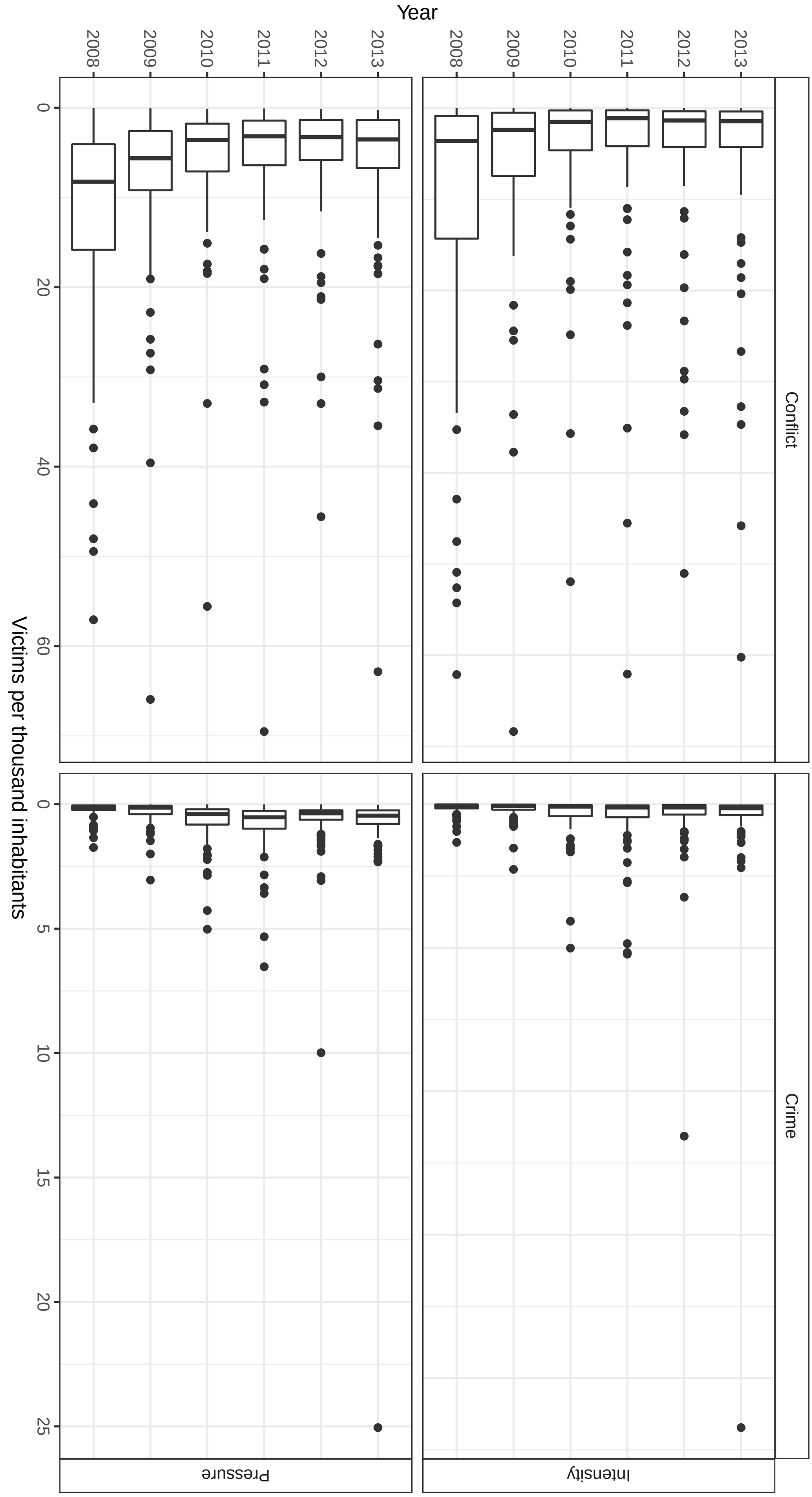

Figure A.13: Distribution of the LA conflict indicators over time

A.3 Appendix to Chapter 5

A.3.1 Linear Regression Model

The linear regression model describes the relationship between a dependent (response) variable y_i and a group of independent (covariate, predictor or regressor) variables as a linear one, which can be written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is the error term, which includes all the random variables that we do not observe but that also affect $\mathbf{y}$. According to the Gauss-Markov Theorem, under certain assumptions, the Ordinary Least Squares (OLS) estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ for the set of parameters $\boldsymbol{\beta}$ is unbiased and efficient. The set of assumptions are:

1. There is a linear relationship between $\mathbf{X}$ and $\mathbf{y}$.
2. The expected value of the random error, conditional on the regressors, is zero; $\mathbf{E}(\boldsymbol{\varepsilon}|\mathbf{X}) = \mathbf{0}$. This implies that any random variable w_i that is not included into $\mathbf{X}$ (as a predictor of $\mathbf{y}$) is uncorrelated with the predictors in $\mathbf{X}$ and is not correlated with the response variable $\mathbf{y}$. This assumption is necessary for the OLS estimators to be unbiased.
3. The error term has conditional variance and covariance matrix $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$. This guarantees that the estimator of the standard error of $\widehat{\boldsymbol{\beta}}$ is unbiased, allowing us to draw inferences from the estimated parameters, and implies two assumptions about the error term $\boldsymbol{\varepsilon}$:
   - Its variance, conditional on $\mathbf{X}$, is the same (σ^2) for all observations. This assumption is called homoscedasticity.
   - The error terms of any pair of observations, i and j are conditionally uncorrelated or $\mathbf{Cov}(\varepsilon_i, \varepsilon_j|\,\mathbf{X}) = 0, \forall i \neq j$.

In this context, in which students are grouped (or nested) into schools, it is very likely that this assumption does not hold, since unobserved variables (such as school institutional environment, for example) may influence the student outcomes. If this is true, the error terms of students A and B in the same school are correlated, making inference invalid. In particular, the information about student A is informative about student B, which implies the sample size is overestimated and thus the standard errors are underestimated.

However, as shown by Hox (2010), this is not the only problem of using the linear regression models for analysing nested data. Using only student-level data risks committing the atomistic fallacy, by which student-level information is used to infer about school-level relationships. On the other side, using only school-level data, implies lower power and risk of committing the ecological fallacy, by which school-level relationships are assumed to hold at the student-level. Using multilevel models instead of single-level regression models can solve all these problems, as the former allows us to simultaneously analyse the interaction of variables that are measured at different levels.

A.3.1.1 Interaction Effects

The discussion in chapter 9 is based on the inclusion of interaction effects between ethnicity and student, school and LA characteristics. This appendix illustrates the two alternative interpretations for such effects using the interaction between SES and ethnicity. This is based on the model

$$maths_i = \beta_0 + \beta_M Minority_i + \beta_{SES} SES_i + \beta_{IE} \left(Minority_i \times SES_i\right) + e_i \tag{A.1}$$

The inclusion of the interaction term implies that the achievement gap between White and minority students is given by $Gap = \beta_M + \beta_{IE} SES_i$, which increases with increases in SES. Simultaneously, the effect of SES on maths achievement changes by ethnicity; being β_{IE} higher for minority students. In the example presented in Figure A.14, both β_M and β_{IE} are negative, resulting in the patterns illustrated below.

Figure A.14 shows the expected achievement for White and minority students. When $SES = 0$, that is, for students with the average SES, the achievement gap is β_M, but as SES increases, the achievement gap increases by β_{IE} per unit increase in SES (as illustrated by the dashed line). Therefore, when $SES = 3$ the achievement gap is $\beta_M + 3\beta_{IE}$. Simultaneously, the slope of the expected achievement line is β_{IE} higher for White students than for minority students, since the effect of a one-unit increase in SES is β_{IE} higher for White than for minority students.

A.3.2 Multilevel Modelling

Analyses using multilevel models usually start with a variance component model (VCM), which allows us to understand how the variation in the dependent variable y_{ij} is explained by within- and between-school variation. For a continuous variable this model

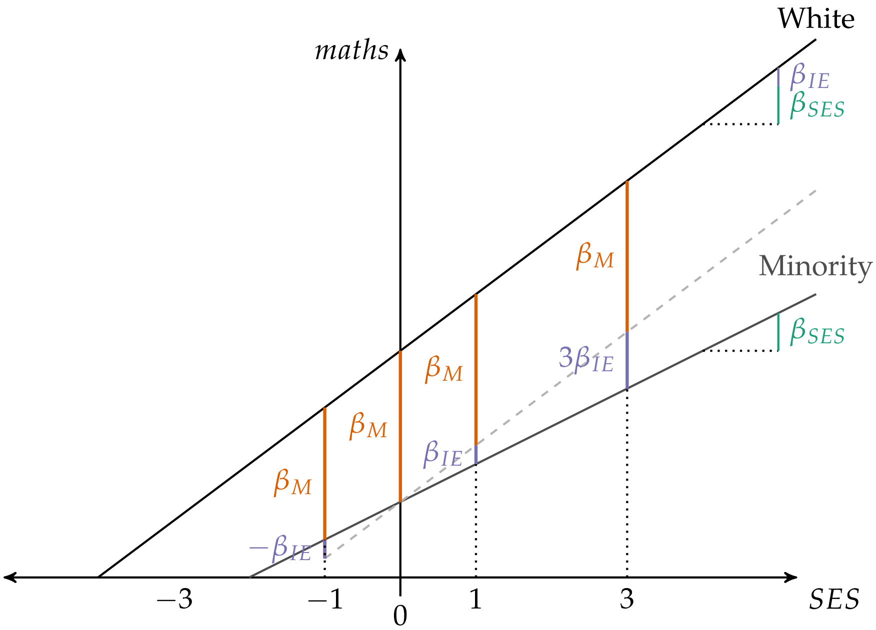

Figure A.14: Illustration of alternative interpretations of interaction terms in a regression model

is

$$
\begin{aligned}
y_{ij} &= \beta_0 + u_{0j} + e_{ij} \\
\\
u_{0j} &\sim N\left(0, \sigma^2_{u_{0_{VCM}}}\right) \\
e_{0j}|u_{0j} &\sim N\left(0, \sigma^2_{e_{VCM}}\right)
\end{aligned}
\tag{A.2}
$$

where y_{ij} is the standardised test score of student i in school j, β_0 is the fixed part of the model, which equals the overall mean test score, and $u_{0j} + e_{ij}$ is the random part of the model or total residual. Here, u_{0j} is the school-level random effect (with variance $\sigma^2_{u_{0_{VCM}}}$), and e_{ij} is the within-school or student-level residual (with variance $\sigma^2_{e_{VCM}}$).

The structure of the random part of the model allows to recognise the correlation of test scores of students within the same school, since all students within school j are affected by u_{0j}. This implies that there is a different mean test score for each school, and that the correlation between two random students in the same school is not assumed to be zero. The mean test score for school j is $\beta_0 + u_{0j}$, which differs from the overall test score by u_{0j}. Additionally, the correlation of test scores for two random students in the same school, known as the intraclass correlation (ICC), is

$$\rho_{VCM} = \frac{\text{Cov}\left(y_{ij}, y_{i'j}\right)}{\sqrt{\text{Var}\left(y_{ij}\right)\text{Var}\left(y_{i'j}\right)}} = \frac{\sigma^2_{u_{0_{VCM}}}}{\sigma^2_{u_{0_{VCM}}} + \sigma^2_{e_{VCM}}}$$

This unconditional intraclass correlation ρ_{VCM}, also represents the proportion of the total variance $\text{Var}\left(y_{ij}\right) = \sigma^2_{u_{0_{VCM}}} + \sigma^2_{e_{VCM}}$ that is explained by between-school variation $\sigma^2_{u_{0_{VCM}}}$.

As with the single-level linear regression model, there are a set of assumptions that are required to obtain consistent and asymptotically efficient estimates of the parameters of the model and consistent estimation of the standard errors through maximum likelihood estimation. This assumptions are:

1. The total residual has mean zero; $\mathbf{E}\left(u_{0j} + e_{ij}\right) = 0$: This implies that the parameter β_0 is the mean of the dependent variable y_{ij} and it is consistently estimated by $\widehat{\beta_0}$. If the total residual follows a symmetric distribution, $\widehat{\beta_0}$ is also an unbiased estimator of β_0. In unbalanced cases, that is, when schools are of different sizes, the maximum likelihood estimator of the overall mean is a precision-weighted average of school means, given by $\widehat{\beta_0} = \frac{\sum_{j=1}^{J} \frac{\bar{y}_{.j}}{\hat{\sigma}^2_{u_{0_{VCM}}} + \frac{\hat{\sigma}^2_{e_{VCM}}}{n_j}}}{\sum_{j=1}^{J} \frac{1}{\hat{\sigma}^2_{u_{0_{VCM}}} + \frac{\hat{\sigma}^2_{e_{VCM}}}{n_j}}}$ (Snijders & Bosker, 2012).

2. There is no correlation between the school- and student-level residuals; $\mathbf{Cov}\left(e_{ij}, u_{0_j}\right) = 0$: This means that the unobserved variables at the student-level are not related to the unobserved variables at the school level. For example, that students' attitudes are not correlated with the schools' institutional environment. This assumption guarantees that the variance of the dependent variable $\mathbf{Var}\left(y_{ij}\right)$ is the addition of the variance of different random effects, that is $\mathbf{Var}\left(y_{ij}\right) = \sigma^2_{u_{0_{VCM}}} + \sigma^2_{e_{VCM}}$.

3. The student-level random effect is not correlated across students i; $\mathbf{Cov}\left(e_{ij}, e_{i'j}\right) = 0$, which implies that the only reason why students in the same school are similar in the unobserved variables is that they attend the same school. That is, $\mathbf{Cov}\left(u_{0j} + e_{ij}, u_{0j} + e_{i'j}\right) = \sigma^2_{u_{0_{VCM}}}$.

4. The random effects in all levels are uncorrelated across different schools; $\mathbf{Cov}\left(e_{ij}, e_{i'j'}\right) = 0$ and $\mathbf{Cov}\left(u_{0j}, u_{0j'}\right) = 0$. This implies that the influence of the unobserved school-level variables are unique for each school and students within that school.

5. When predicting the random effects using Bayes estimators, that the random effect follows a normal distribution.

The inclusion of student- ($\mathbf{x}_{ij}$) and school-level covariates ($\mathbf{w}_j$) is straight-forward, as with the single-level regression model. The model is expressed as

$$\begin{gathered} y_{ij} = \beta_0 + \mathbf{x}'_{ij}\boldsymbol{\beta}_1 + \mathbf{w}'_j\boldsymbol{\beta}_2 + u_{0j} + e_{ij} \\ \\ u_{0j}|\mathbf{w}_j \sim N\left(0, \sigma^2_{u_0}\right) \\ e_{0j}|\mathbf{w}_j, u_{0j} \sim N\left(0, \sigma^2_e\right) \end{gathered} \tag{A.3}$$

Consistent estimators for the regression parameters can be obtained under the following assumptions:

1. The functional form an the covariates are correctly specified.

2. Level-1 exogeneity; $\mathbf{E}\left(e_{ij}|\mathbf{x}_{ij}, \mathbf{w}_j, u_{0j}\right) = 0$: This implies that the student-level error is neither correlated with the regressors nor with the random intercept.

3. Level-2 exogeneity; $\mathbf{E}\left(u_{0j}|\mathbf{x}_{ij}, \mathbf{w}_j\right) = 0$: This implies that the school-level random part is not correlated with the covariates.

Additionally, for the standard errors of the estimated parameters to be consistent, it is assumed:

1. Homoscedastic student-level residuals, given the predictors and school-level random effect; $\mathbf{Var}\left(e_{ij}|\mathbf{x}_{ij}, \mathbf{w}_j, u_{0j}\right) = \sigma^2_e$: The variance of the student-level residual, conditional on the random school intercept and school-level covariates, is the same for all students.

2. Homoscedastic school-level residuals, given the predictors; $\mathbf{Var}\left(u_{0j}|\mathbf{w}_j\right) = \sigma^2_{u_0}$: The variance of the random school-intercept, conditional on the regressors, is the same for all schools.

3. Uncorrelated student-level residuals, given the predictors and school-level random effect; $\mathbf{Cov}\left(e_{ij}, e_{i'j'}|\mathbf{x}_{ij}, \mathbf{x}_{i'j'}, \mathbf{w}_j, \mathbf{w}_{j'}, u_{0j}, u_{0j'}\right) = 0, \forall i \neq i', \forall j \neq j'$: Even if students are within the same school, the student-level error is not correlated among students, after taking into account the covariates and the school-level random effects.

4. Uncorrelated school-level residuals, given the predictors; $\mathbf{Cov}\left(u_{0j}, u_{0j'}|\mathbf{w}_j, \mathbf{w}_{j'}\right) = 0$: The school-level random effects of two different schools are not correlated, conditional on the regressors.

All these assumptions are required for efficient estimation of the parameters, and a symmetric distribution of the random part of the model is also required for unbiased estimation.

In order to understand in what proportion the covariates (in this case, $\mathbf{x}_{ij}$ and $\mathbf{w}_j$) explain the between- and within-school y_{ij} variation, it is possible to compute the proportions in which these variance components reduce after including the covariates, which are

$$R^2_{u_0} = \frac{\sigma^2_{u_{0_{VCM}}} - \sigma^2_{u_0}}{\sigma^2_{u_{0_{VCM}}}}$$

for the between-school test score variance component, and

$$R^2_e = \frac{\sigma^2_{e_{VCM}} - \sigma^2_e}{\sigma^2_{e_{VCM}}}$$

for the within-school test score variance component. The proportion of the variance that is explained by all covariates is given by an overall R^2, which is

$$R^2 = \frac{\sigma^2_{u_{0_{VCM}}} + \sigma^2_{e_{VCM}} - \sigma^2_{u_0} - \sigma^2_e}{\sigma^2_{u_{0_{VCM}}} + \sigma^2_{e_{VCM}}}$$

which is basically the reduction in the total variance of y_{ij} that is explained by the inclusion of the covariates $\mathbf{x}_{ij}$ and $\mathbf{w}_j$. However, since the inclusion of student-level variables can reduce both the student- and the school-level residual variance of y_{ij}, both $R^2_{u_0}$ and R^2 can be negative.

A.3.3 Generalised Version of a Mediation Analysis Model

The basic mediation model includes three regression equations, represented by the path diagram in Figure A.15. The upper part of this figure represents the total effect of the independent variable x_i on the dependent variable y_i, β_t, which is equivalent to the regression model $y_i = \alpha_t + \beta_t x_i + e_{t_i}$. The bottom part of Figure A.15 shows that the relationship between x_i and y_i is mediated by m_i. The regression equations corresponding to this part are

$$m_i = \alpha_m + \beta_{xm} x_i + e_{m_i} \tag{A.4}$$

which estimates the overall effect of x_i on the mediator variable m_i, β_{xm} and

$$y_i = \alpha_y + \beta_{xy} x_i + \beta_{my} m_i + e_{y_i} \tag{A.5}$$

which estimates the effect of x_i on y_i, conditional on m_i, β_{xy}.

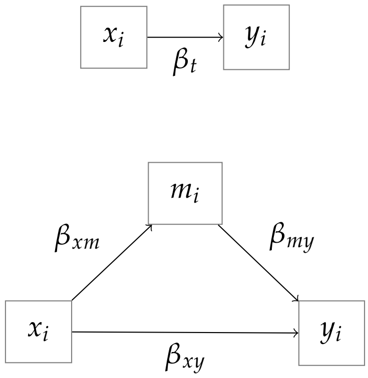

Figure A.15: Path diagram for the basic mediation model

In this model, the total effect of x_i on y_i, β_t is split into its direct effect β_{xy} and its indirect or mediated effect $\beta_{xm}\beta_{my}$ or, equivalently, $\beta_t - \beta xy$ (MacKinnon et al., 2007). This leads to the relationship

$$\beta_t = \beta_{xy} + \beta_{xm}\beta_{my} \tag{A.6}$$

A.3.4 Software

Basic data preparation used Stata 15 (StataCorp, 2017), while the analysis was conducted using R 3.3.3 (R Core Team, 2017). Multilevel estimation was performed in MLwiN 3.03 using the R2MLwiN package (Charlton, Rasbash, Browne, Healy, & Cameron, 2019; Z. Zhang, Parker, Charlton, Leckie, & Browne, 2016). The figures were created using the ggplot2 package (Wickham, 2016). Finally, this document was created using knitr, which combines the functionality of R and LaTeX (Xie, 2019).

A.3.5 Ethical Considerations

This research adheres to the ethical principles proposed by the Economic and Social Research Council (ESRC) Framework for Research Ethics (ESRC, 2015) (as this council funds this research), the University of Bristol Research Governance and Integrity Policy

(University of Bristol, 2016), and the British Education Research Association (BERA) Ethical Guidelines for Educational Research (BERA, 2011). Furthermore, this research received the approval of the University of Bristol Graduate School of Education Ethics Committee on March 2nd, 2017, under the application ID 49942.

Regarding the protection of research participants, this research uses administrative data (collected by Icfes) that is publicly available, and does not involve direct participation of students nor school or LA representatives. Students were asked to confirm they authorised Icfes to make their personal data available for research through the registration forms for the exam since 2009. In the SABER 11 database, Students are identified by a number supplied by Icfes, and there is no contact information. Schools can be identified by name and a unique identification number from the DANE. Since schools tend to be small and to have a low proportion of minority students, it may be possible to identify individual students from the dataset. However, the risks entailed by this possibility are minimum because the databases date from 2013 or earlier and registered students that were about to finish their last year of school.

Additionally, none of the research questions require the disclosure of specific students or schools, but to point out general patterns of the educational system. The policies of both Icfes and Economic and Social Research Council (ESRC) require researchers to make their research data available for the public. In order to minimise the risk of disclosing the identity of any individual student, schools will be given an arbitrary identification number that cannot be matched with other databases that identify schools, and the ethnicity of students will be only disclosed as general ethnic groups (Afrocolombian, Indigenous or other) and not to the specific Indigenous or Afrocolombian group they belong.

Also abiding by the principles in ESRC (2015), the methods that are used to analyse the data have been carefully selected, in such a way that they are appropriate to answer the research questions. Even more, the analysis of how the substantive research results are affected by different methods is a central aspect of this research. This is the case because of the potential implications for educational and social policy in Colombia. For this reason, the limitations of the research and the need to conduct further research to answer specific questions before outlining any specific policy recommendation are clearly discussed.

Finally, the influence of the funding body (the ESRC) and the University of Bristol on the research is limited to the provision of a set of standards and guidelines for research quality. By no means the results that are presented or the way they are interpreted are influenced by particular interests of these or any other institutions or particulars, including the researcher

A.4 Appendix to Chapter 6

A.4.1 Achievement Gap Decomposition for Multiple Groups and Levels

A generalised version of the model that allows decomposing the overall gap of G groups in within- and between- L levels gaps is shown in figures A.16 and A.17.

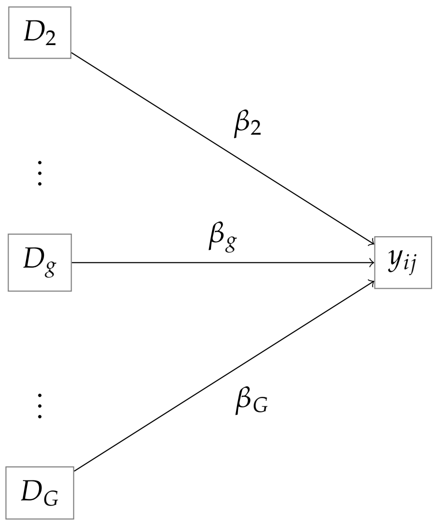

Figure A.16: Generalised total effects model for G groups and L levels

In the general case, the categories of the independent variable (groups) are indexed by $g = 1, \ldots, G$ (where 1 is taken as the reference group), and the levels are indexed by $l = 1, \ldots, L$. The $ijk \ldots$ subscripts have been omitted to avoid confusion. The variables y and D_g always vary by individual (level 1) and the mediators vary at level l. The subscripts of the coefficients indicate the variable they multiply within each equation. The superscripts indicate the level and the equation to which the coefficient belongs.

The overall gap is given by

$$y = \alpha + \sum_{g=2}^{G} \beta_g D_g + e_y \tag{A.7}$$

The outcomes model is equivalent to the contextual effects model

$$y = \alpha + \sum_{g=2}^{G} \beta_g^W D_g + \sum_{l=1}^{L} \sum_{g=2}^{G} \beta^{lg} \, \mathrm{E}\left[D_g | l = l\right] + e_y \tag{A.8}$$

Therefore, the $(G-1) \times L$ mediation equations are

$$\begin{array}{rcl}
\mathrm{E}\left[D_2|l=1\right] & = & \omega_{012} + \sum_{g'=2}^{G} \omega_{g'}^{12} D_{g'} + e_{12} \\
& \vdots & \\
\mathrm{E}\left[D_G|l=1\right] & = & \omega_{01G} + \sum_{g'=2}^{G} \omega_{g'}^{1G} D_{g'} + e_{1G} \\
& \vdots & \\
\mathrm{E}\left[D_g|l=l\right] & = & \omega_{0lg} + \sum_{g'=2}^{G} \omega_{g'}^{lg} D_{g'} + e_{lg} \\
& \vdots & \\
\mathrm{E}\left[D_G|l=L\right] & = & \omega_{0LG} + \sum_{g'=2}^{G} \omega_{g'}^{LG} D_{g'} + e_{LG}
\end{array} \tag{A.9}$$

and the $G-1$ gaps can be decomposed into the within- and between-levels as

$$\beta_g = \left(1 - \omega_g^{1g}\right) \beta_g^W + \sum_{l=2}^{L} \left(\omega_g^{(l-1)g} - \omega_g^{lg}\right) \beta^{B(l-1)g} + \omega_g^{Lg} \beta^{BLg} + \sum_{l=1}^{L} \sum_{g' \neq g} \omega_g^{lg'} \beta^{lg'} \tag{A.10}$$

where $\beta^{Blg} = \beta_g^W + \sum_{l=1}^{L} \beta^{lg}$.

One example of this general formulation is the decomposition of the gaps of the different minority groups at the within-school, between-school and LA levels.

The outcomes model is the contextual effect model (8.1), given by

$$y_{ijk} = \alpha + \beta_A^W A_{ijk} + \beta_I^W I_{ijk} + \beta_O^W O_{ijk} + \beta^{CSA} \overline{A}_{.jk} + \beta^{CSI} \overline{I}_{.jk} + \beta^{CSO} \overline{O}_{.jk} + \\ \beta^{CLA} \overline{A}_{..k} + \beta^{CLI} \overline{I}_{..k} + \beta^{CLO} \overline{O}_{..k} + e_{y_{ijk}} \tag{A.11}$$

The six mediation models are

$$\begin{array}{rcl}
\overline{A}_{.jk} & = & \gamma_1 + \omega_A^{SA} A_{ijk} + \omega_I^{SA} I_{ijk} + \omega_O^{SA} O_{ijk} + e_{SA.jk} \\
\overline{I}_{.jk} & = & \gamma_2 + \omega_A^{SI} A_{ijk} + \omega_I^{SI} I_{ijk} + \omega_O^{SI} O_{ijk} + e_{SI.jk} \\
\overline{O}_{.jk} & = & \gamma_3 + \omega_A^{SO} A_{ijk} + \omega_I^{SO} I_{ijk} + \omega_O^{SO} O_{ijk} + e_{SO.jk} \\
\overline{A}_{..k} & = & \gamma_4 + \omega_A^{LA} A_{ijk} + \omega_I^{LA} I_{ijk} + \omega_O^{LA} O_{ijk} + e_{LA..k} \\
\overline{I}_{..k} & = & \gamma_5 + \omega_A^{LI} A_{ijk} + \omega_I^{LI} I_{ijk} + \omega_O^{LI} O_{ijk} + e_{LI..k} \\
\overline{O}_{..k} & = & \gamma_6 + \omega_A^{LO} A_{ijk} + \omega_I^{LO} I_{ijk} + \omega_O^{LO} O_{ijk} + e_{LO..k}
\end{array} \tag{A.12}$$

Therefore the decomposition is given by

$$\beta_A = \left(1 - \omega_A^{SA}\right) \beta_A^W + \left(\omega_A^{SA} - \omega_A^{LA}\right) \beta^{BSA} + \omega_A^{LA} \beta^{BLA} + \\ \omega_A^{SI} \beta^{CSI} + \omega_A^{SO} \beta^{CSO} + \omega_A^{LI} \beta^{CLI} + \omega_A^{LO} \beta^{CLO}$$

$$\beta_I = \left(1 - \omega_I^{SI}\right)\beta_I^W + \left(\omega_I^{SI} - \omega_I^{LI}\right)\beta^{BSI} + \omega_I^{LI}\beta^{BLI} + \\ \omega_I^{SA}\beta CSA + \omega_I^{SO}\beta^{CSO} + \omega_I^{LA}\beta^{CLA} + \omega_I^{LO}\beta^{CLO} \quad \text{(A.13)}$$

$$\beta_O = \left(1 - \omega_O^{SO}\right)\beta_O^W + \left(\omega_O^{SO} - \omega_O^{LO}\right)\beta^{BSO} + \omega_O^{LO}\beta^{BLO} + \\ \omega_O^{SA}\beta^{CSA} + \omega_O^{SI}\beta^{CSI} + \omega_O^{LA}\beta^{CLA} + \omega_O^{LI}\beta^{CLI}$$

where $\beta^{BLg} = \beta_g^W + \beta^{CSg} + \beta^{CLg}$.

A.4.2 Results for Heterogeneous Schools

This subsection reproduces the figures of subsection 6.5 with a restricted sample, which excludes schools that serve only one ethnic group. The next tables show how many students and schools are excluded from the sample using this criterion.

Table A.10: Number of students in ethnically homogeneous schools by ethnic group and year

	2008	2009	2010	2011	2012	2013
White	197,171	217,518	221,015	232,997	236,845	160,739
Afro	1,887	1,543	1,514	2,205	2,772	7,171
Indigenous	1,015	784	805	1,157	1,550	3,317
Other	305	206	238	435	445	2,847

Table A.11: Number of schools in ethnically homogeneous schools by ethnic group and year

	2008	2009	2010	2011	2012	2013
White	4,696	456	284	274	215	174
Afro	84	12	8	6	6	12
Indigenous	39	5	5	7	11	13
Other	19	0	3	4	2	6

A.4.3 Estimation Results for Schools that are Observed During all Years in the Sample

This subsection reproduces the figures of subsection 6.5 with a sample consisting of schools that are observed during all years in the sample.

Table A.12: Number of students in schools observed all years by ethnic group and year

	2008	2009	2010	2011	2012	2013
White	331,729.00	361,650.00	381,935.00	373,032.00	356,433.00	348,676.00
Afro	10,488.00	11,781.00	12,304.00	13,021.00	13,533.00	18,947.00
Indigenous	5,146.00	5,521.00	5,998.00	6,265.00	6,249.00	9,858.00
Other	6,174.00	5,071.00	6,324.00	6,615.00	5,260.00	22,165.00

Table A.13: Proportion of schools observed all years by year

	2008	2009	2010	2011	2012	2013
Schools	10.40	6.77	4.27	3.95	3.29	3.04

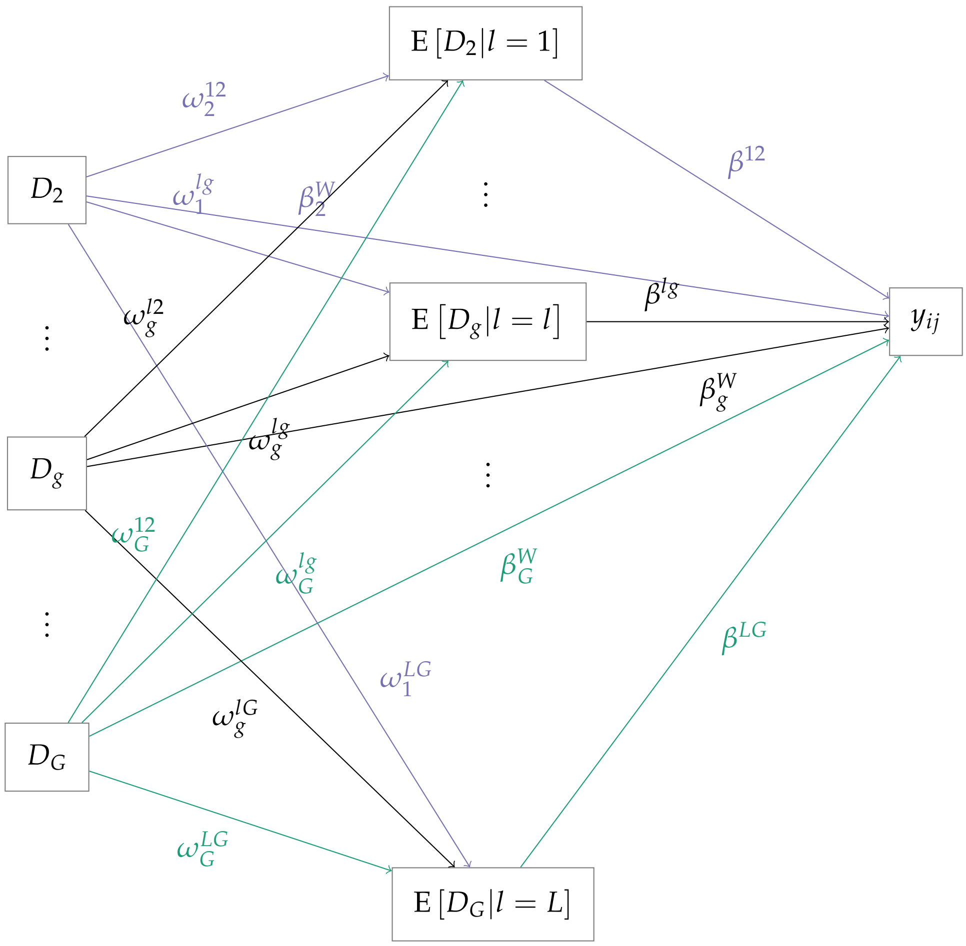

Figure A.17: Generalised outcome and mediation models for G groups and L levels

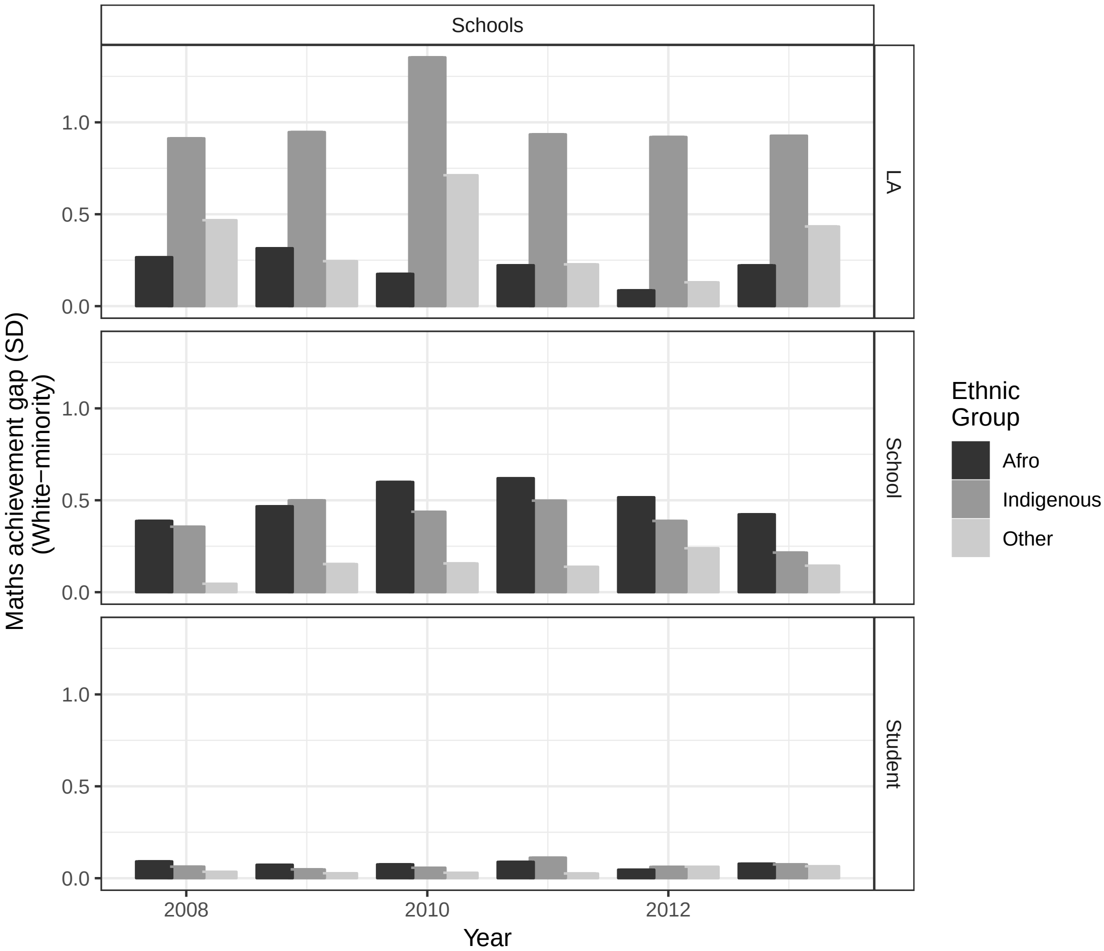

Figure A.18: Within-school gaps and school and LA contextual effect of each minority group for the heterogeneous-school sample

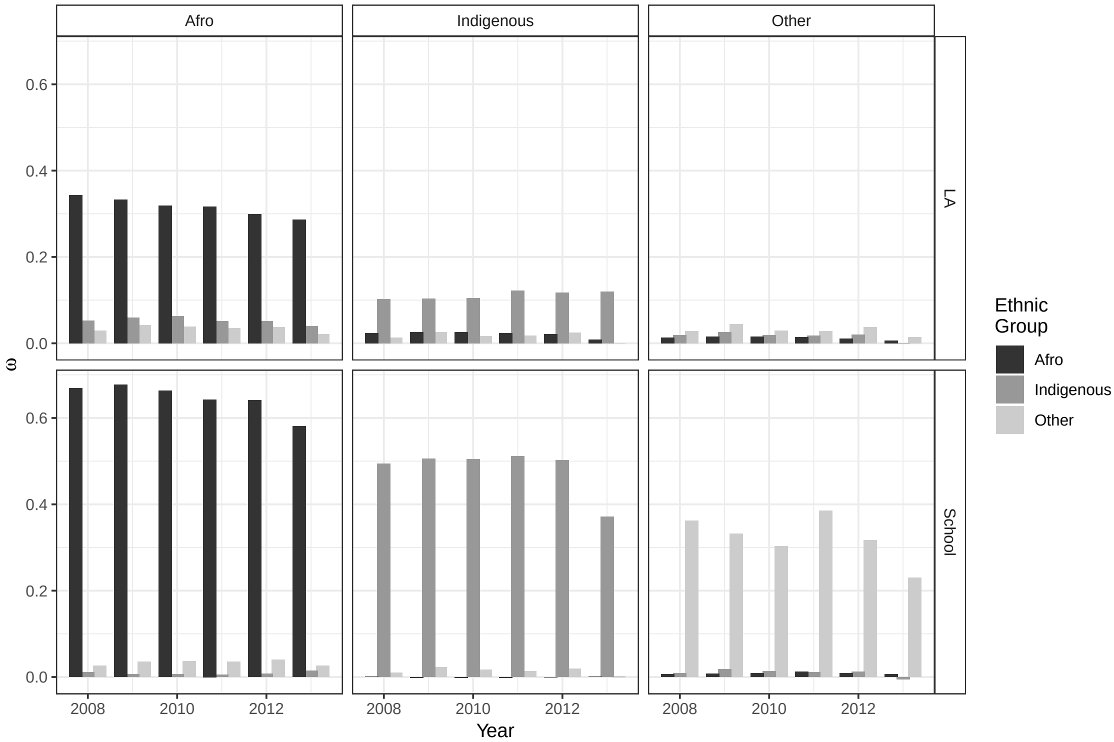

Figure A.19: Segregation indices ω_Y^X in the achievement gap decomposition for the heterogeneous-school sample

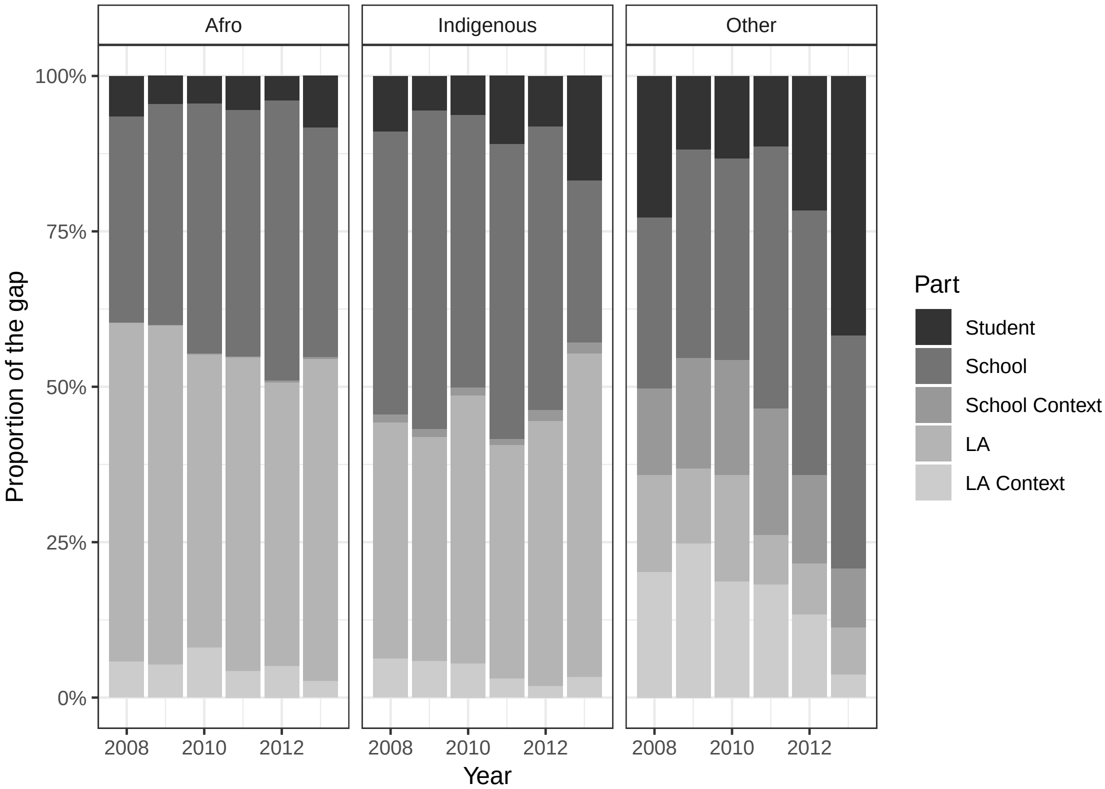

Figure A.20: Decomposition of the overall achievement gap between White and Afro-colombian, Indigenous and other minority students in heterogeneous schools

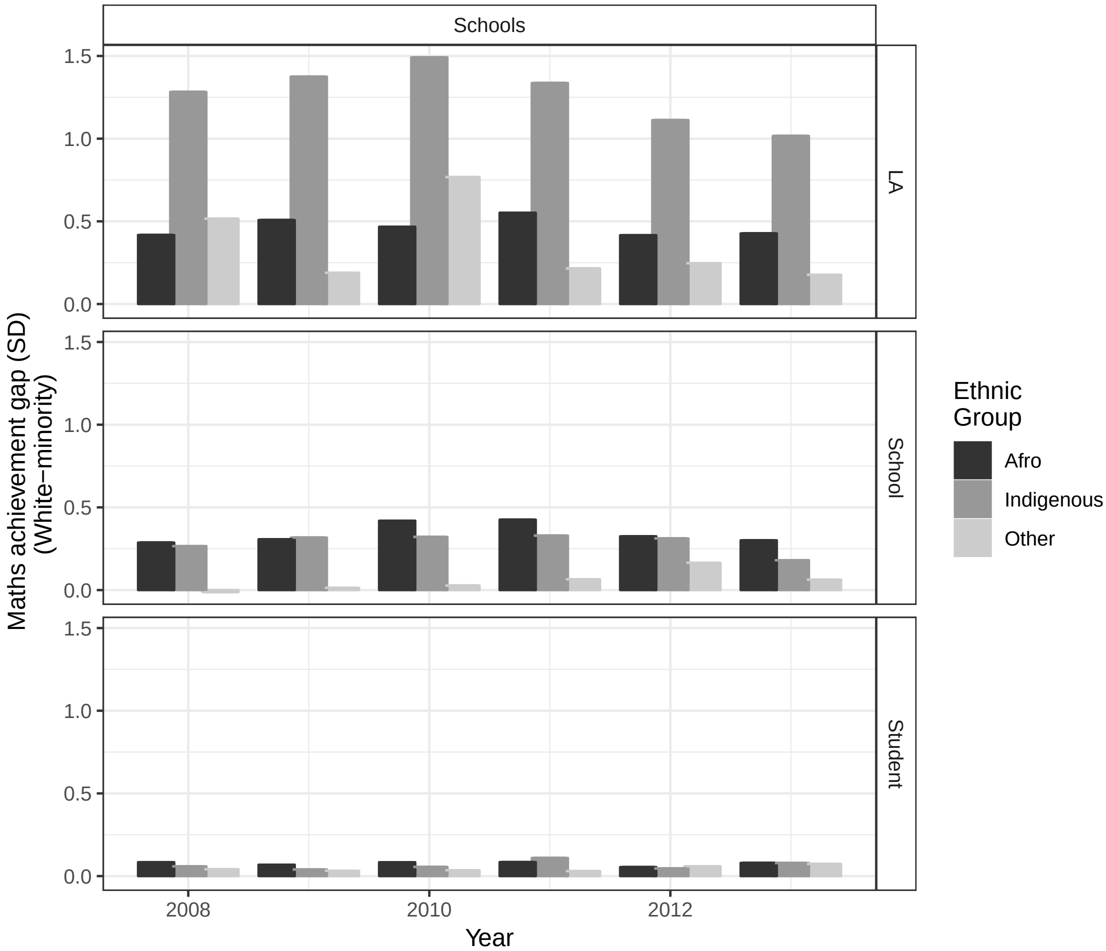

Figure A.21: Within-school gaps and school and LA contextual effect of each minority group for the sample of schools observed during all years

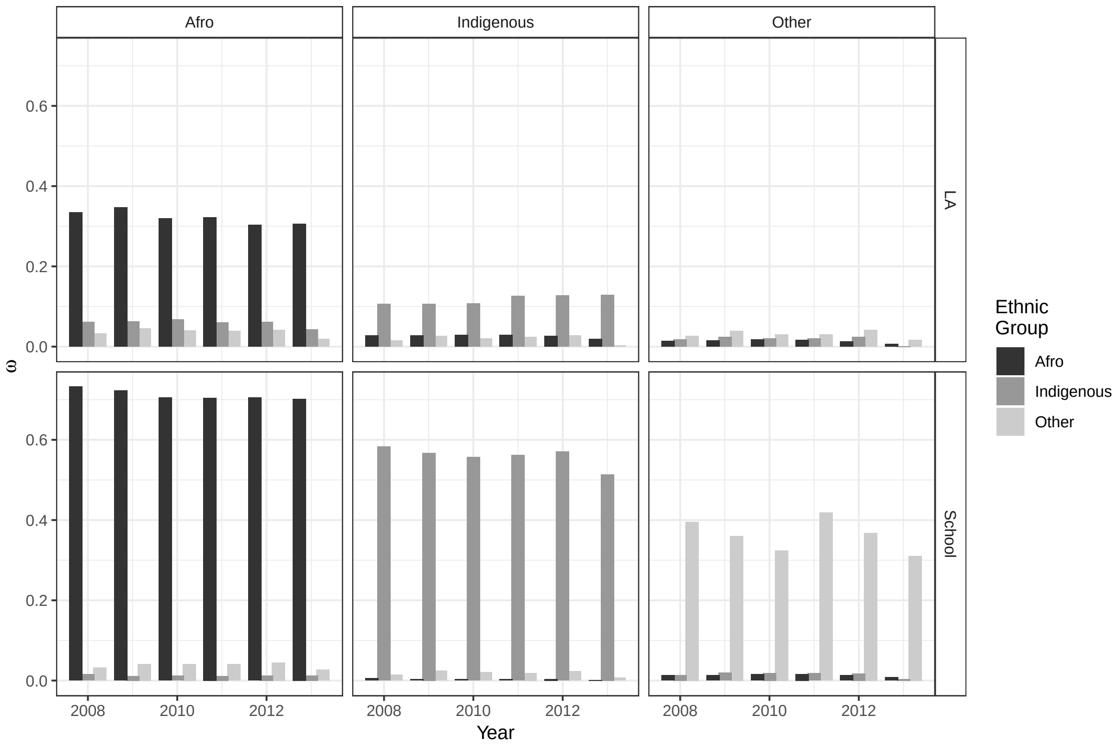

Figure A.22: Segregation indices ω_Y^X in the achievement gap decomposition for the sample of schools observed during all years

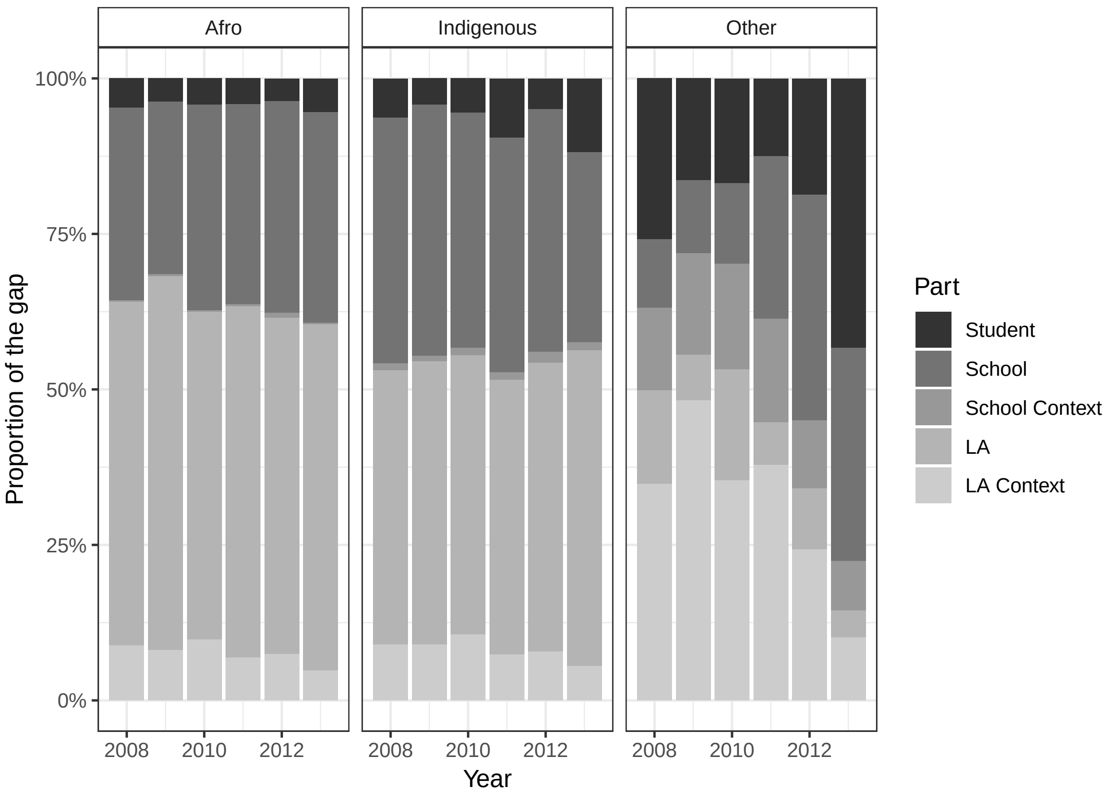

Figure A.23: Decomposition of the overall achievement gap between White and Afro-colombian, Indigenous and other minority students in the sample of schools observed during all years

A.5 Appendix to chapter 7

A.5.1 Principal Components Analysis and Factor Analysis

Principal component analysis (PCA) and Factor analysis (FA) are two of the index construction techniques that are analysed in chapter 7, which focuses on its application and how it compares to other techniques that have been used in the ethnic achievement gap literature. This subsection briefly describes these methods.

A.5.1.1 Principal Components Analysis

Principal component analysis (PCA) is a data reduction technique that transforms a set of correlated continuous variables (x_i, $i = 1, \ldots, p$, such as education, occupation and income) into its principal components (pc_j, $j = 1, \ldots, p$). The first principal component pc_1 explains the largest proportion of the total variance and each consecutive component explains a smaller proportion of the variance. Each principal component pc_j is a weighted sum of the original variables x_i, with weights a_{ij} that preserve the original ordering of the observations. Finding these weights is equivalent to finding the eigenvalues and eigenvectors of the data matrix.

Some principal components are positively correlated with all the xs, so they are interpreted as a measure of what the variables have in common. For example, in chapter 7, where PCA is used to create a SES index using education, occupation and income as the set of variables, it is assumed that the first component represents the students' SES. In other applications (such as questionnaire design) additional components may also be of interest and they are interpreted according to what is common to the different subsets of variables with larger weights.

To construct an indicator, the original variables are combined using a standardised version of the weights a_{ij} of the first principal component, given by $\tilde{pc}_j = \sum_{i=1}^{p} \frac{a_{ij}}{\sqrt{\lambda_j}} x_i$, where $x_i, i = 1, \ldots, p$ have been standardised to have mean zero and unit variance and λ_j is the variance of the j-th principal component. Each principal component pc_j explains a proportion of the total variance of the original variables x_i. Therefore, a measure of fit for the first principal component is the proportion of explained variance $\frac{\lambda_1}{\sum_{j=1}^{p} \lambda_j}$.

When PCA is used to gain information about a latent variable, the aim is to include the components that explain the largest proportion of the total variance. There are different criteria to decide how many components to retain:

- Retain the first k components until explaining 70-90% of the total variance.

- When using standardised variables (and thus, analysing the correlation matrix), retain components with an eigenvalue of at least 0.7.

- Choose the k first components until the eigenvalues start decreasing more slowly.

- Include the components with a sensible and useful interpretation.

To determine what variables have a larger contribution to the component j, the component loadings $a*_{ij} = \sqrt{\lambda_j a_{ij}}$ are analysed. If the variables have been standardised, these are directly interpretable. The communality for variable i is useful to understand how well the component scores represent the original set of variables x_i, which is $\sum_{j=1}^{k} \lambda a_{ij}^2$. The closest to one, the better variable x_i is represented by the first k components.

Table A.14: Principal components rotation and standard deviations

	PC 1	PC 2	PC 3	PC 4	PC 5
s. father_ed cont.	0.510	-0.260	0.319	-0.333	-0.678
s. mother_ed cont.	0.518	-0.328	0.044	-0.348	0.707
s. father_oc cont.	0.333	0.891	0.270	-0.117	0.094
s. mother_oc cont.	0.386	0.142	-0.894	-0.024	-0.174
s. income cont.	0.459	-0.107	0.152	0.868	0.032
Standard Deviation	1.585	0.909	0.865	0.767	0.570

PC: Principal Component

A.5.1.2 Factor Analysis

Factor analysis (FA) aims to understand whether the correlation between observed variables x_i, $i = 1, \ldots, p$ can be explained by the interaction of one or a small number of latent (unobserved) variables, called factors $f_1, \ldots, f_q$. This technique [6] is based on a

[6] FA is part of a the family of latent variable models, which differ on the type of observed variables $x_1, \ldots, x_p$ and factors that they include, generating different techniques. If both, the factors and the observed variables are metrical, the method is called FA; if the factors are metrical but the observed variables are categorical, the procedure is known as latent trait analysis (LTA); if it is the opposite way (categorical factor and metrical observed variables), the technique is latent profile analysis; finally, if both, factors and observed variables are categorical, the name of the process is latent class analysis (LCA). As shown in chapter 7, the most common technique in the literature is FA.

model in which the observable variables are explained by the same set of factors, given by

$$x_i = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} f_j + e_i \tag{A.14}$$

where the α_{ij}s are known as factor loadings.

In the case of the creation of a SES index, the observed variables x_i are education, occupation and income, and the factor represents the SES. Unlike PCA, FA incorporates the uncertainty in the relationship between the observed and latent variables, represented by $e_i \sim N\left(0, \sigma_i^2\right)$. Model A.14 is thus estimated using iterative maximum likelihood, assuming that the factors f_j have mean 0 and variance of 1 and are uncorrelated with each other and with the error terms e_i, called uniqueness in this context. As with PCA, a measure of fit of the estimated factor is the proportion of explained variance.

The construction of the SES indicator is more complex in the context of FA than for PCA, as it involves not only considering the observed variables x_i but also the uniqueness of each variable e_i. In chapter 7, the SES index is computed as the factor scores estimated using Bartlett (1937)'s method, which produces unbiased estimates of the factor scores (and therefore of SES) (Skrondal & Laake, 2001).

Table A.15: Factor analysis loadings and uniquenesses

	Loadings	Uniqueness
s. father_ed cont.	0.79	0.377
s. mother_ed cont.	0.816	0.334
s. father_oc cont.	0.367	0.865
s. mother_oc cont.	0.457	0.791
s. income cont.	0.594	0.647
χ^2		20,073

A.5.2 Additional Tables and Figures

Table A.16: Descriptive statistics for the continuous variables and indices by ethnic group

	Overall			White			Afro			Indigenous			Other		
	mean	S.D.	median	mean	S.D.	median	mean	S.D.	median	mean	S.D.	median	mean	S.D.	median
father_ed cont.	7.41	5.34	5.00	7.48	5.33	5.00	6.32	5.19	5.00	4.78	4.86	5.00	7.54	5.48	5.00
mother_ed cont.	7.61	5.14	5.00	7.69	5.12	5.00	6.48	5.19	5.00	4.66	4.89	5.00	7.61	5.31	5.00
father_oc cont.	41.67	16.02	38.00	41.79	16.08	38.00	39.02	14.29	38.00	39.05	13.69	38.00	41.61	16.96	38.00
mother_oc cont.	19.73	25.03	0.00	19.88	25.11	0.00	18.20	23.56	0.00	12.24	20.94	0.00	20.67	25.43	0.00
income cont.	2.03	1.99	1.50	2.07	2.01	1.50	1.33	1.16	1.50	1.01	0.95	0.50	2.01	2.07	1.50
FA	0.00	1.10	-0.14	0.02	1.10	-0.13	-0.30	1.02	-0.53	-0.68	0.93	-0.75	0.01	1.14	-0.10
PCA	0.00	1.59	-0.31	0.03	1.59	-0.26	-0.46	1.37	-0.66	-0.96	1.24	-1.23	0.02	1.63	-0.20
s. average	-0.00	0.70	-0.14	0.01	0.70	-0.11	-0.20	0.60	-0.32	-0.41	0.54	-0.55	0.01	0.72	-0.10
s. sum	-0.00	3.51	-0.69	0.06	3.52	-0.57	-1.00	3.01	-1.58	-2.04	2.71	-2.75	0.05	3.62	-0.50
sum	78.45	38.94	70.50	78.92	39.06	71.50	71.35	35.34	60.50	61.74	31.55	49.50	79.45	39.78	74.50
Hollingshead	33.79	15.02	30.00	33.97	15.06	30.00	30.58	13.62	27.00	27.87	12.34	24.00	34.39	15.41	30.00
prediction cat.	0.04	0.33	-0.05	0.05	0.33	-0.04	-0.07	0.25	-0.14	-0.16	0.21	-0.22	0.04	0.34	-0.05
prediction cont.	0.51	0.32	0.45	0.52	0.32	0.46	0.40	0.23	0.36	0.31	0.21	0.26	0.51	0.33	0.46

S.D: Standard deviation. For the definition of the variables please refer to Table 7.2.

Table A.17: Estimation model for the creation of a composite indicator using prediction from categorical variables

	maths
Intercept	-0.04^{***}
	(0.005)
Father-incomplete primary	0.444^{***}
	(0.011)
Father-primary	0.124^{***}
	(0.01)
Father-incomplete secondary	0.042^{***}
	(0.009)
Father-secondary	0.019^{**}
	(0.008)
Father-incomplete vocational	0.073^{***}
	(0.006)
Father-vocational	0.038^{***}
	(0.007)
Father-incomplete professional	0.047^{***}
	(0.008)
Father-professional	-0.018^{**}
	(0.007)
Father-postgraduate	0.001
	(0.009)
Mother-incomplete primary	0.435^{***}
	(0.011)
Mother-primary	-0.003
	(0.01)
Mother-incomplete secondary	0.015
	(0.009)
Mother-secondary	0.006
	(0.008)

Continued on next page

Table A.17 – *Continued*

	maths
Mother-incomplete vocational	0.027***
	(0.006)
Mother-vocational	0.009
	(0.007)
Mother-incomplete professional	0.015**
	(0.007)
Mother-professional	−0.009
	(0.007)
Mother-postgraduate	−0.003
	(0.008)
Father-pensioner	0.085***
	(0.01)
Father-others	0.026**
	(0.011)
Father-laborer	0.053***
	(0.01)
Father-self-employed worker	−0.05***
	(0.009)
Father-manager	−0.005
	(0.008)
Father-general manager	−0.077***
	(0.008)
Father-technician	0.024***
	(0.007)
Father-secretary	−0.037***
	(0.007)
Father-self-employed professional	0.008
	(0.007)
Father-small entrepreneur	0.014**
	(0.006)

Continued on next page

Table A.17 – *Continued*

	maths
Father-entrepreneur	0.027***
	(0.008)
Mother-pensioner	0.031***
	(0.011)
Mother-others	−0.019
	(0.011)
Mother-laborer	−0.033***
	(0.009)
Mother-self-employed worker	−0.003
	(0.011)
Mother-manager	−0.012
	(0.011)
Mother-general manager	−0.007
	(0.011)
Mother-technician	0.051***
	(0.01)
Mother-secretary	−0.026***
	(0.01)
Mother-self-employed professional	0.016
	(0.009)
Mother-small entrepreneur	−0.034***
	(0.008)
Mother-entrepreneur	0.038***
	(0.011)
Income-[1,2)	0.11***
	(0.004)
Income-[2,3)	0.211***
	(0.005)
Income-[3,5)	0.336***
	(0.007)

Continued on next page

Table A.17 – *Continued*

	maths
Income-[5,7)	0.498***
	(0.011)
Income-[7,10)	0.651***
	(0.014)
Income-10 or more	0.927***
	(0.013)
R^2	0.107
adj.-R^2	0.107
Observations	383,875
$\widehat{\sigma}$	0.945

Standard errors in parenthesis.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

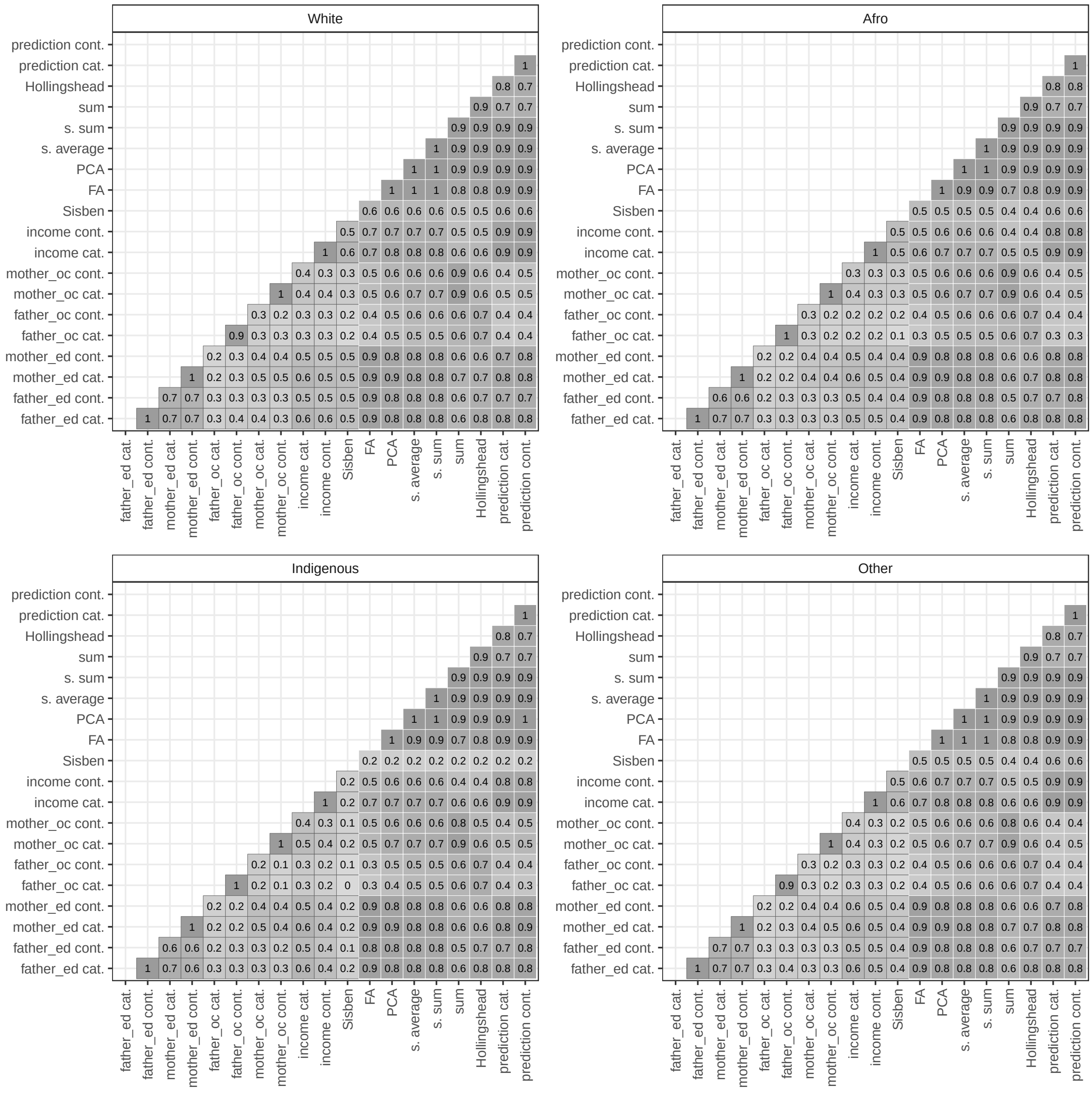

Figure A.24: Correlations (to the nearest 0.1) between observed variables and SES composite indicators for each ethnic group

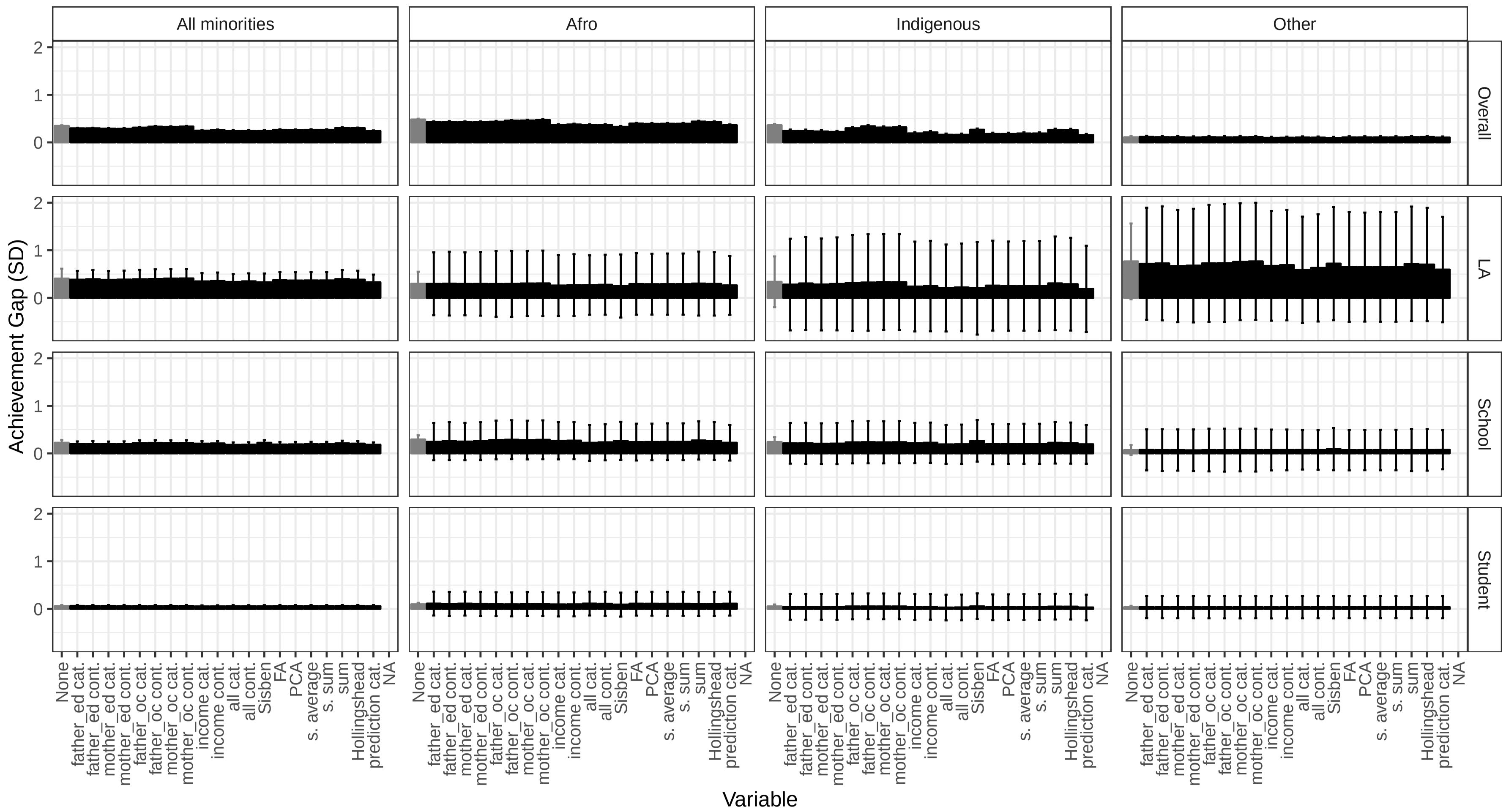

Figure A.25: Estimated conditional ethnic maths achievement gaps by SES composite indicator

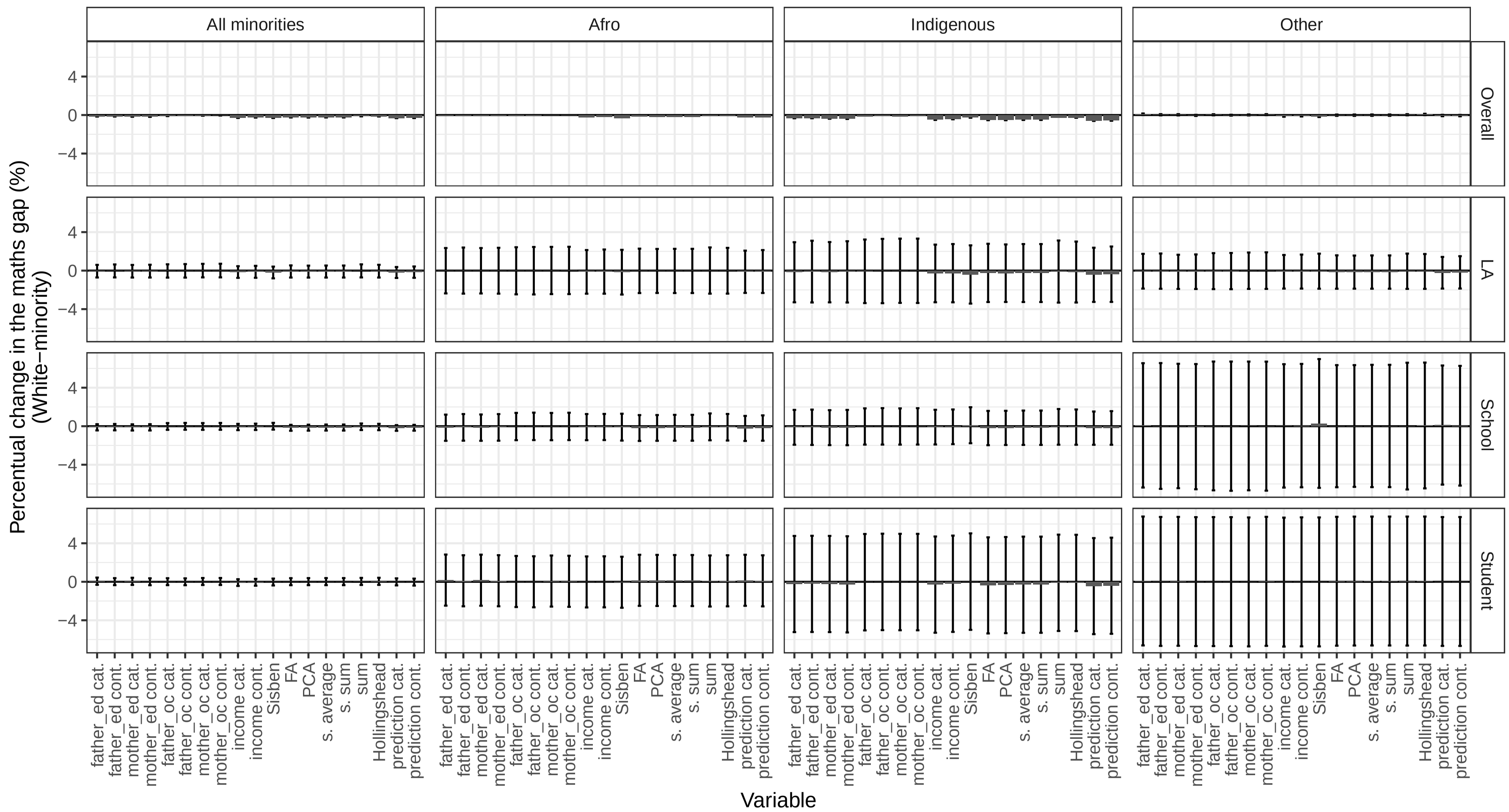

Figure A.26: Confidence intervals for the percentage change in the estimated conditional ethnic maths achievement gaps by SES composite indicator

A.6 Appendix to Chapter 8

A.6.1 Results for Heterogeneous Schools

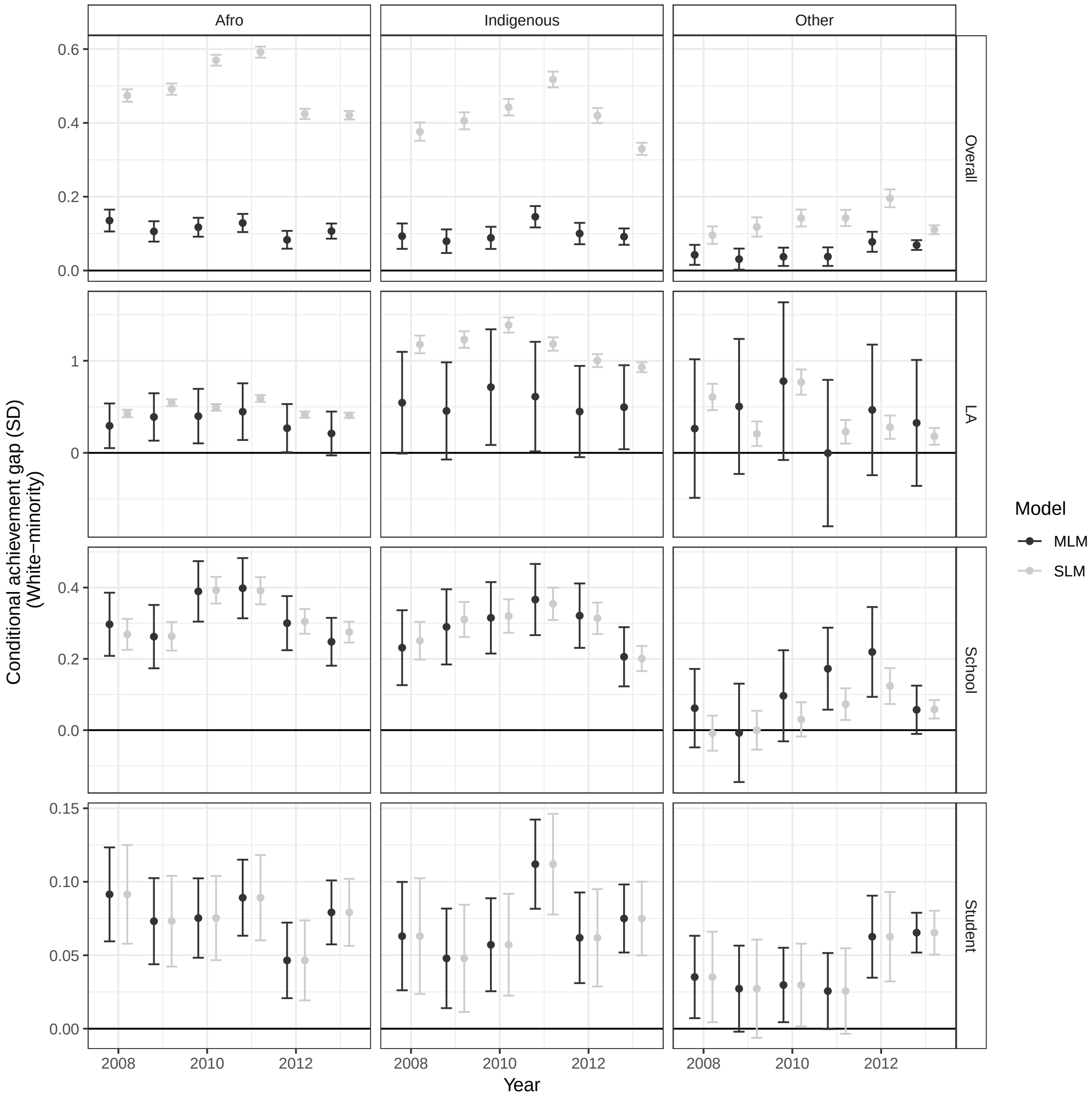

Figure A.27: Estimated unconditional and conditional overall ethnic achievement gaps and their components by estimation method for the sample of heterogeneous schools

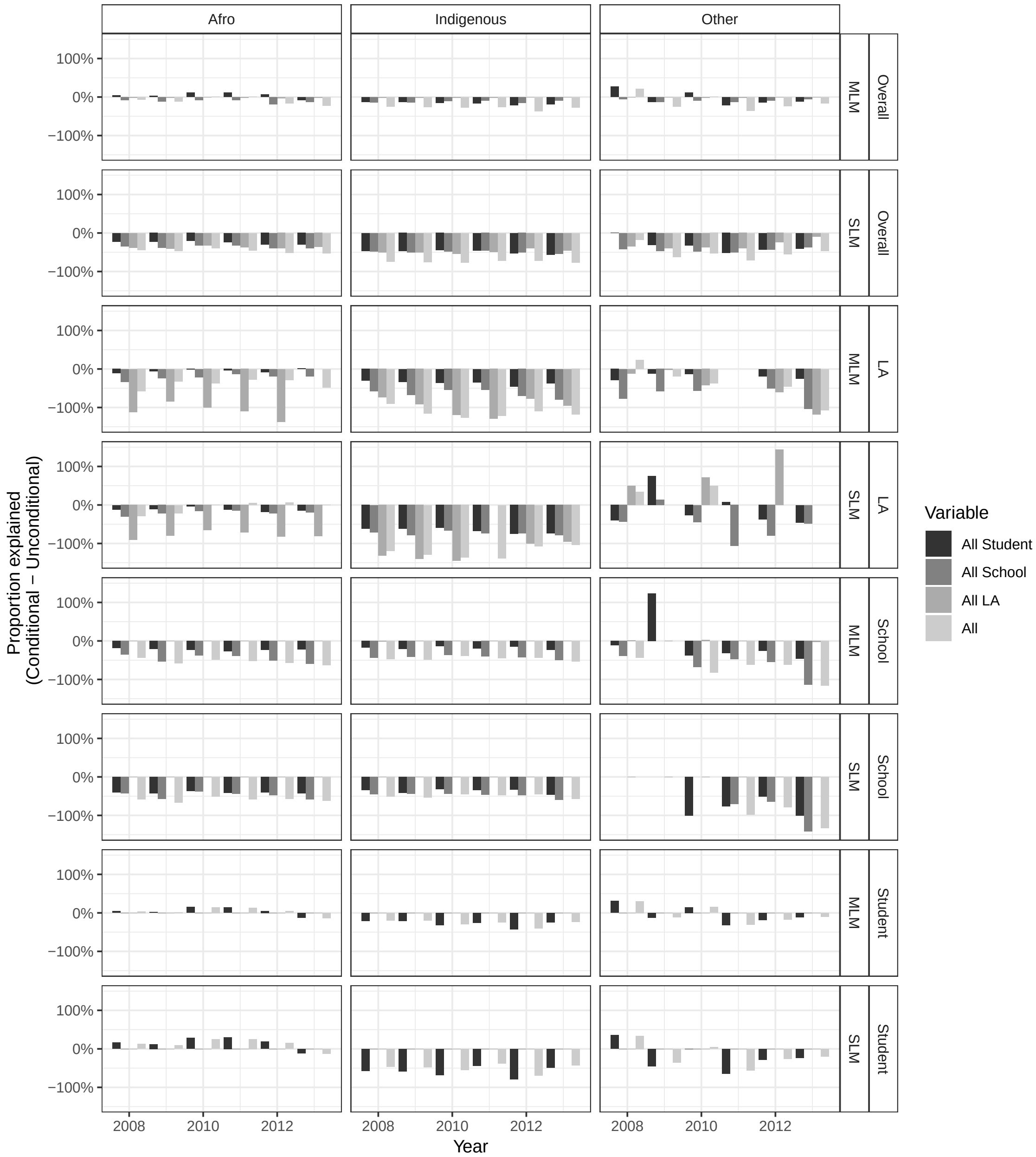

Figure A.28: Estimated changes in the overall gaps and their components after controlling for student, school and LA characteristics by estimation method and set of control variables for the sample of heterogeneous schools

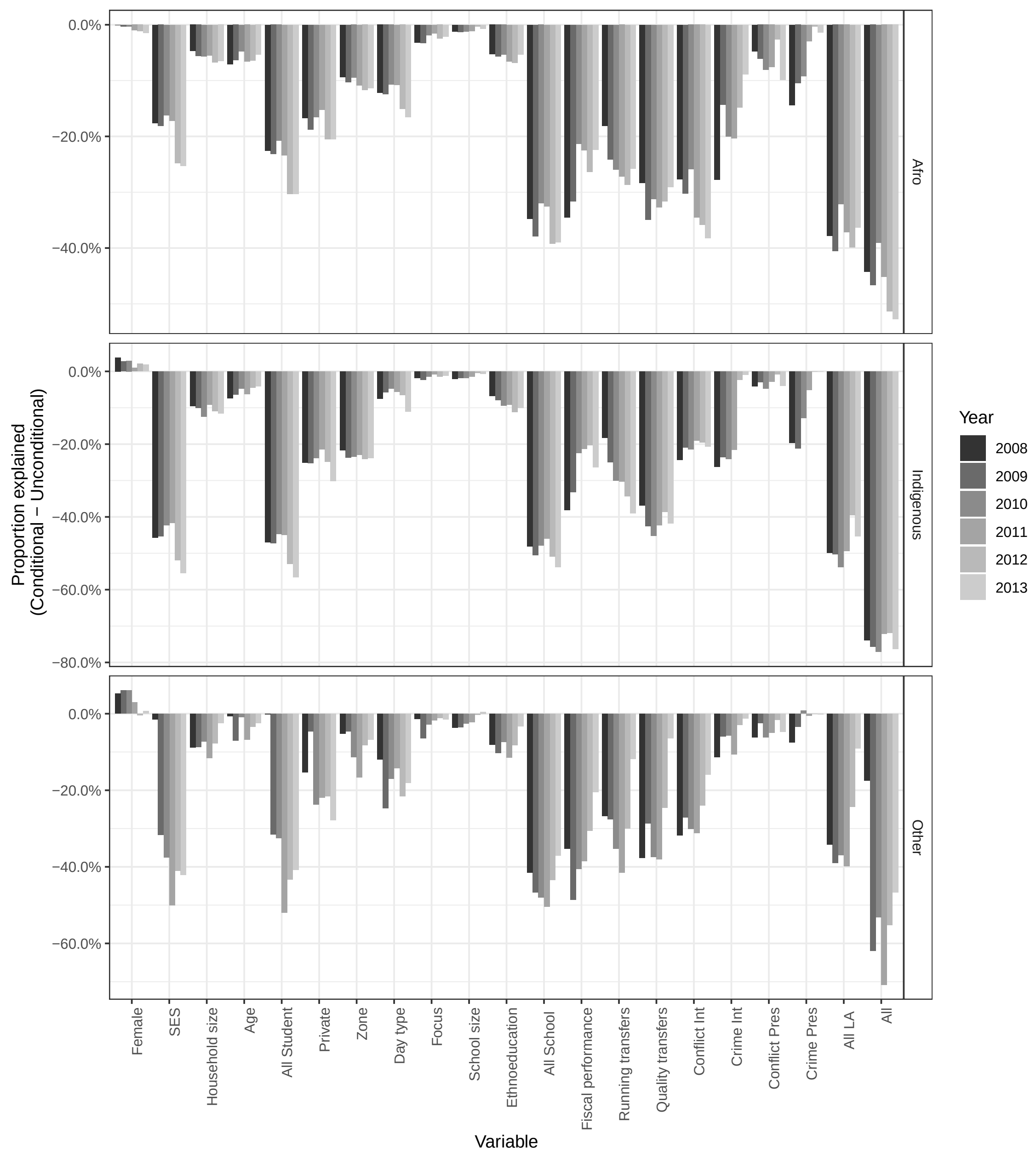

Figure A.29: Estimated effect of student, school and LA characteristics on the overall ethnic achievement gaps for the sample of heterogeneous schools

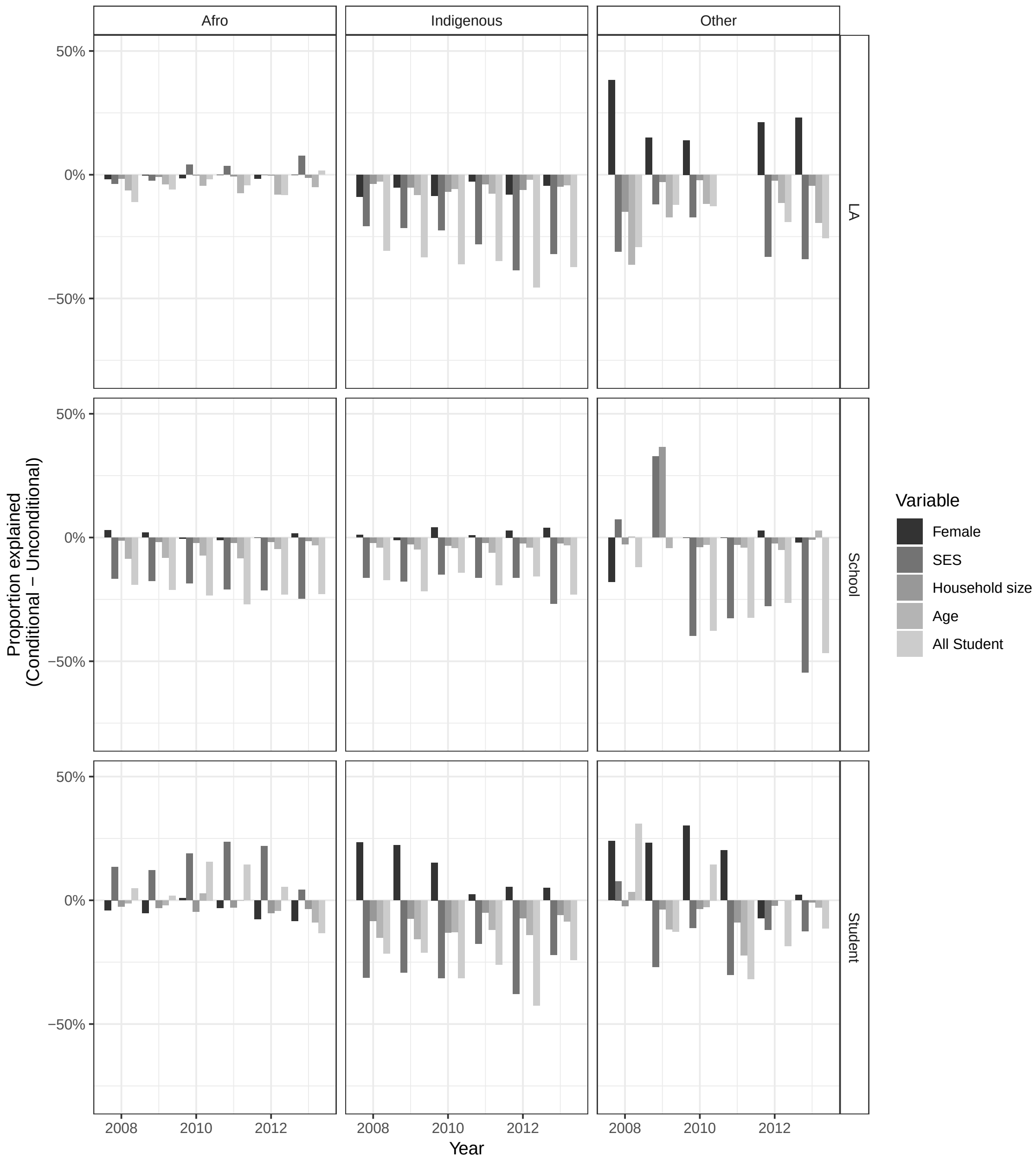

Figure A.30: Estimated effect of student characteristics on the within-school ethnic achievement gaps and the school and LA contextual effect of ethnicity for the sample of heterogeneous schools

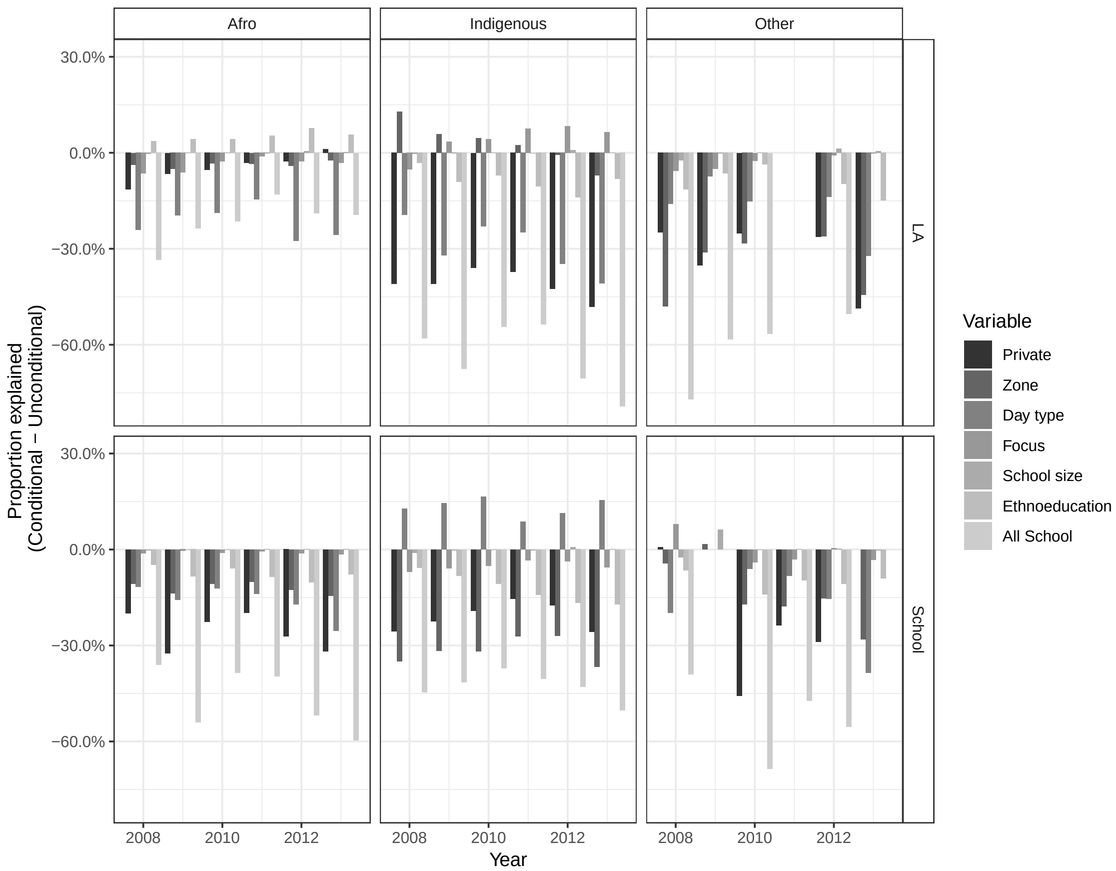

Figure A.31: Estimated changes in the school and LA contextual effects of ethnicity by school characteristic for the sample of heterogeneous schools

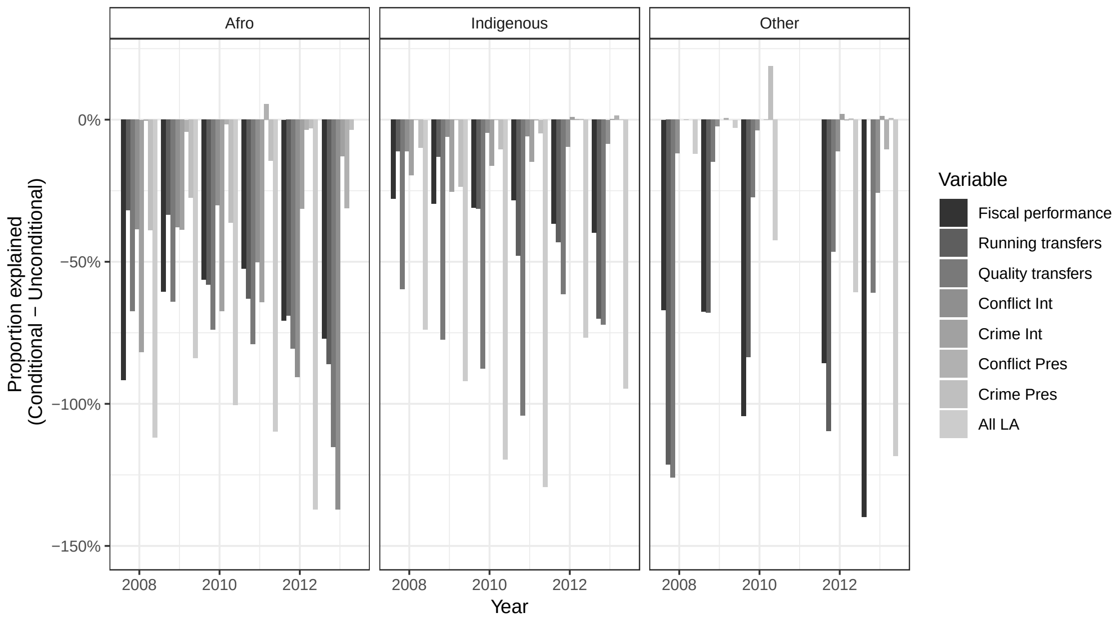

Figure A.32: Estimated changes in the LA contextual effect of ethnicity by LA characteristic for the sample of heterogeneous schools

A.6.2 Results for Schools that are Observed During all Years in the Sample

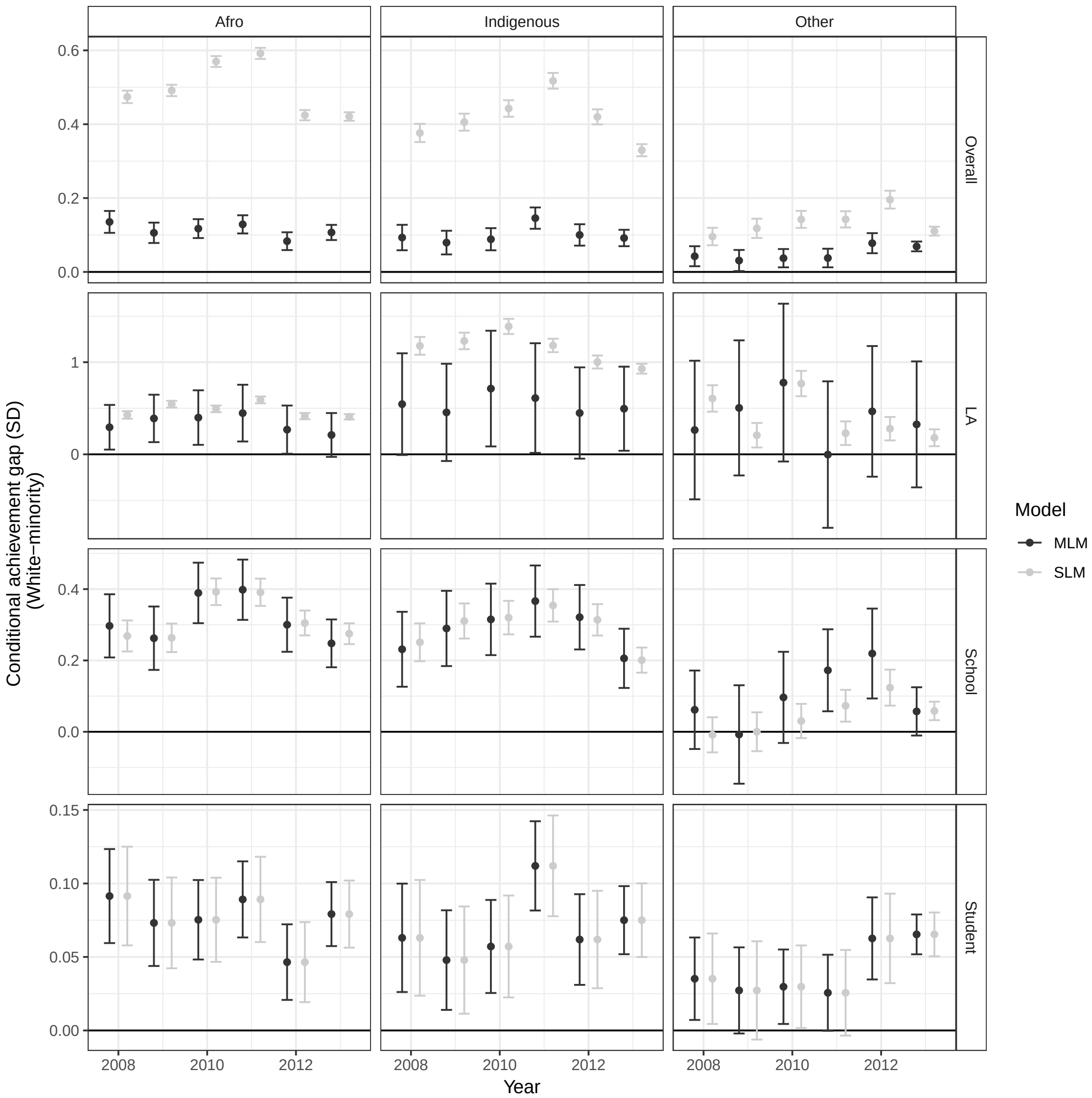

Figure A.33: Estimated unconditional and conditional overall ethnic achievement gaps and their components by estimation method for the sample of schools that are observed during all years

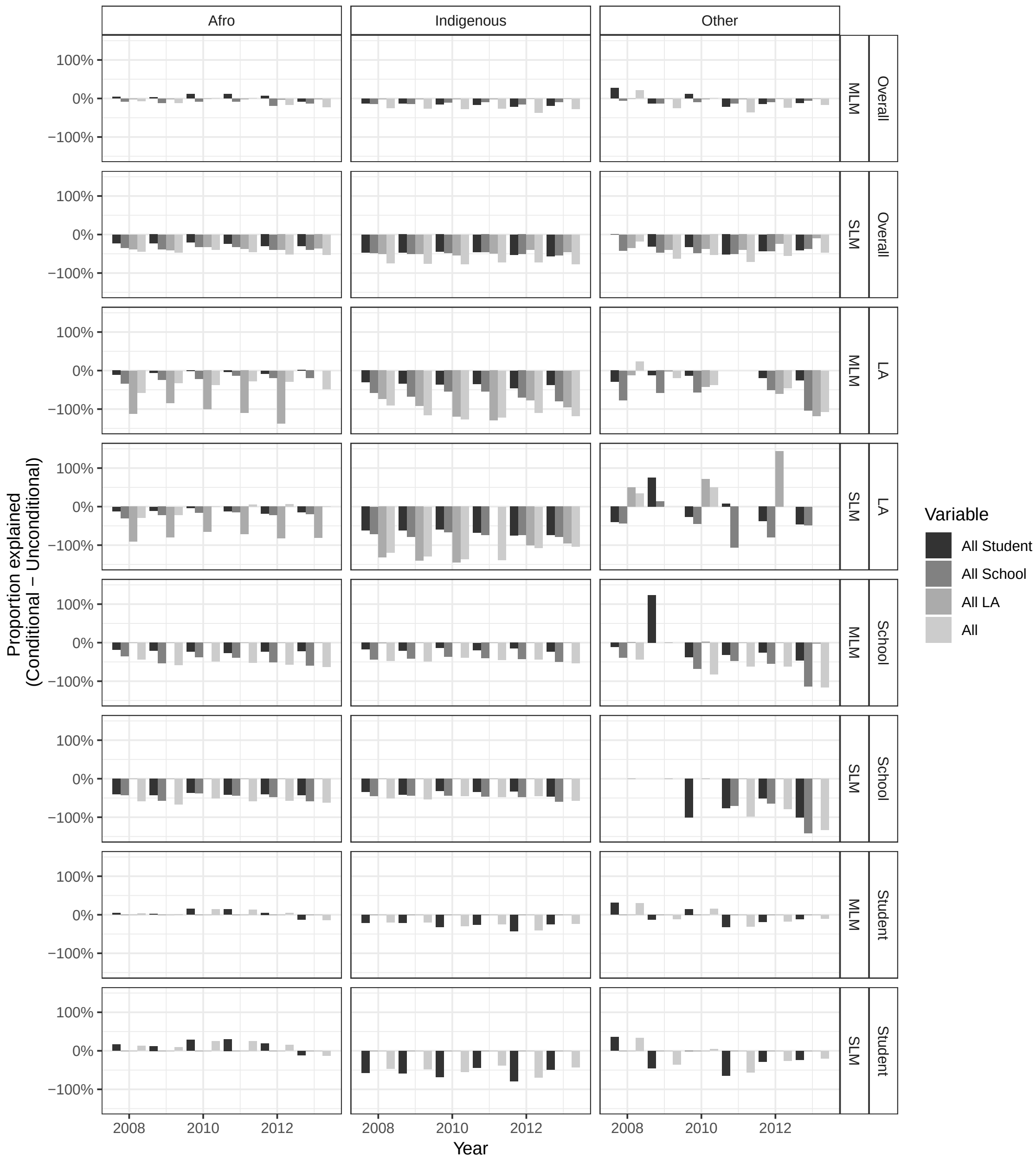

Figure A.34: Estimated changes in the overall gaps and their components after controlling for student, school and LA characteristics by estimation method and set of control variables for the sample of schools that are observed during all years

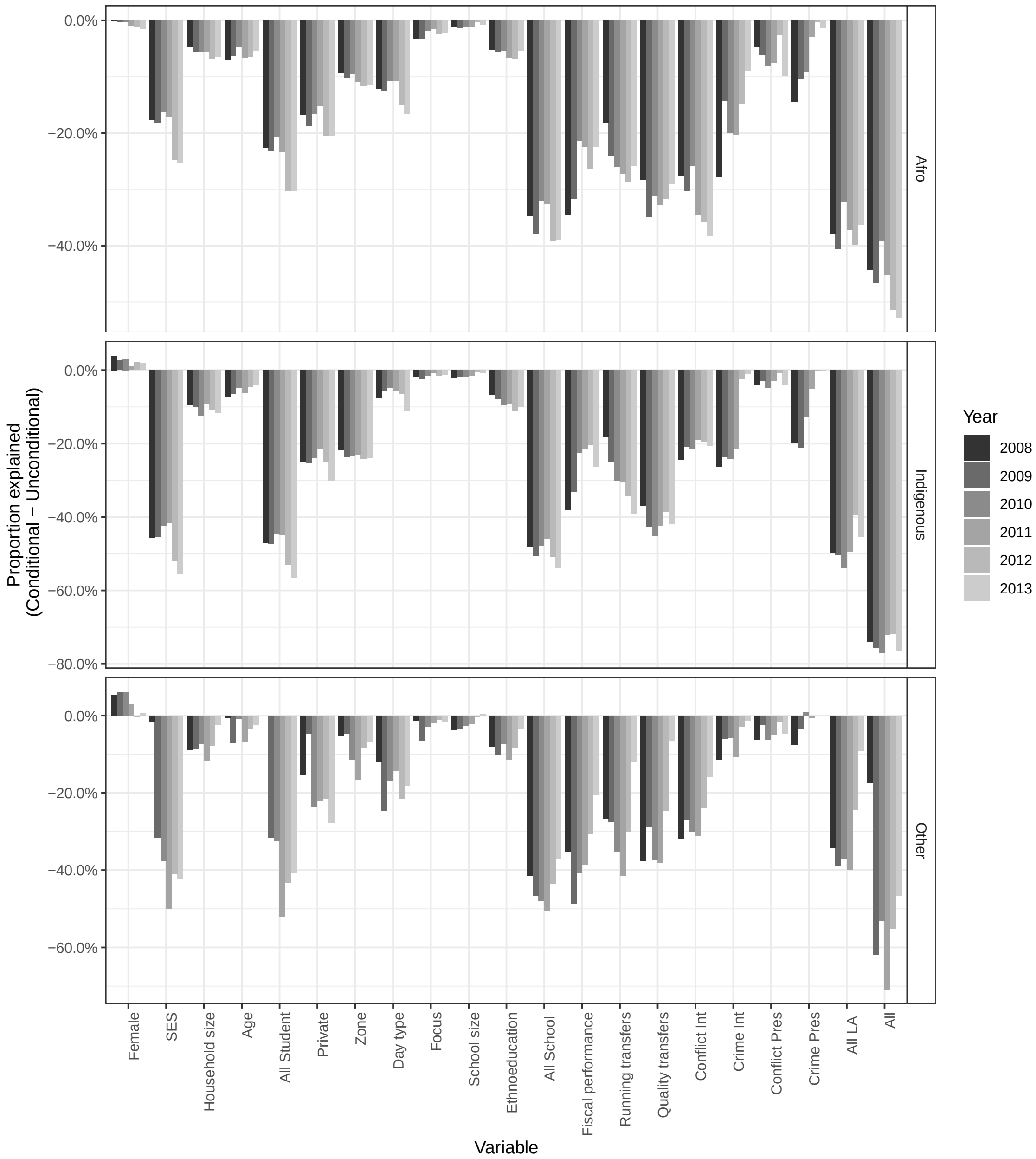

Figure A.35: Estimated effect of student, school and LA characteristics on the overall ethnic achievement gaps for the sample of schools that are observed during all years

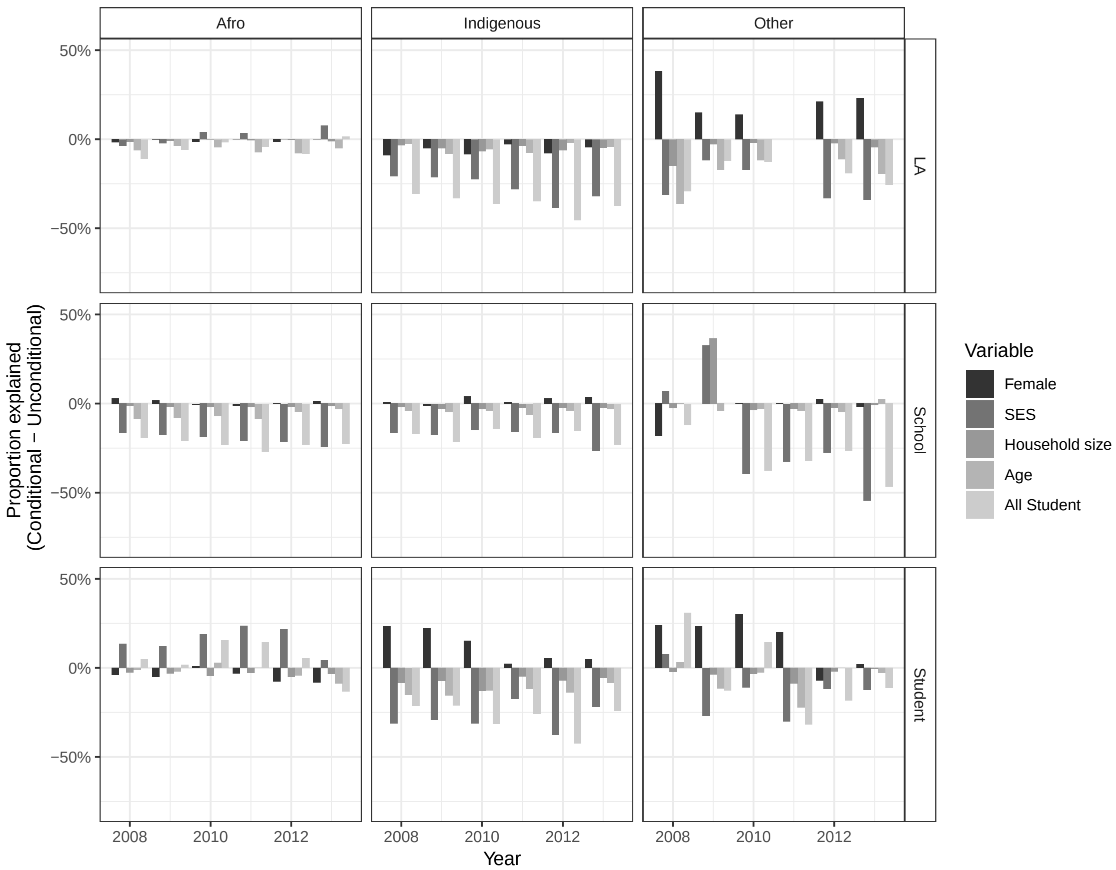

Figure A.36: Estimated effect of student characteristics on the within-school ethnic achievement gaps and the school and LA contextual effect of ethnicity for the sample of schools that are observed during all years

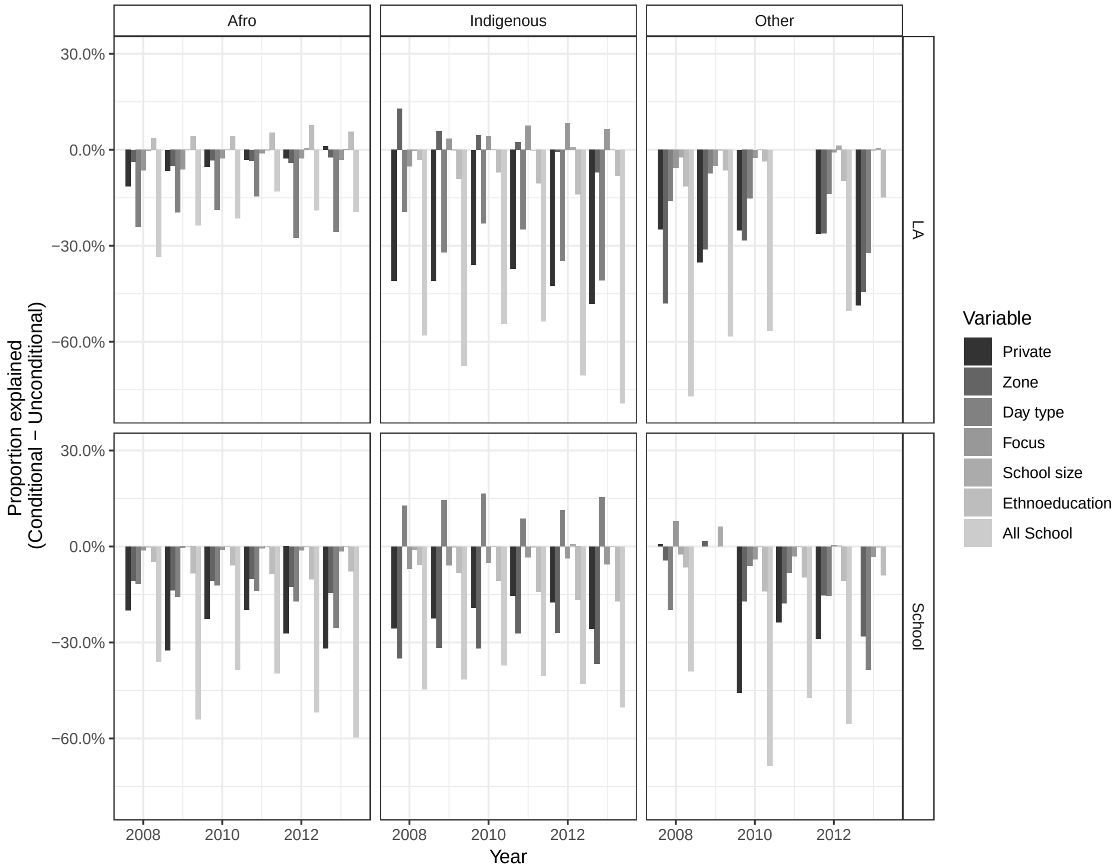

Figure A.37: Estimated changes in the school and LA contextual effects of ethnicity by school characteristic for the sample of schools that are observed during all years

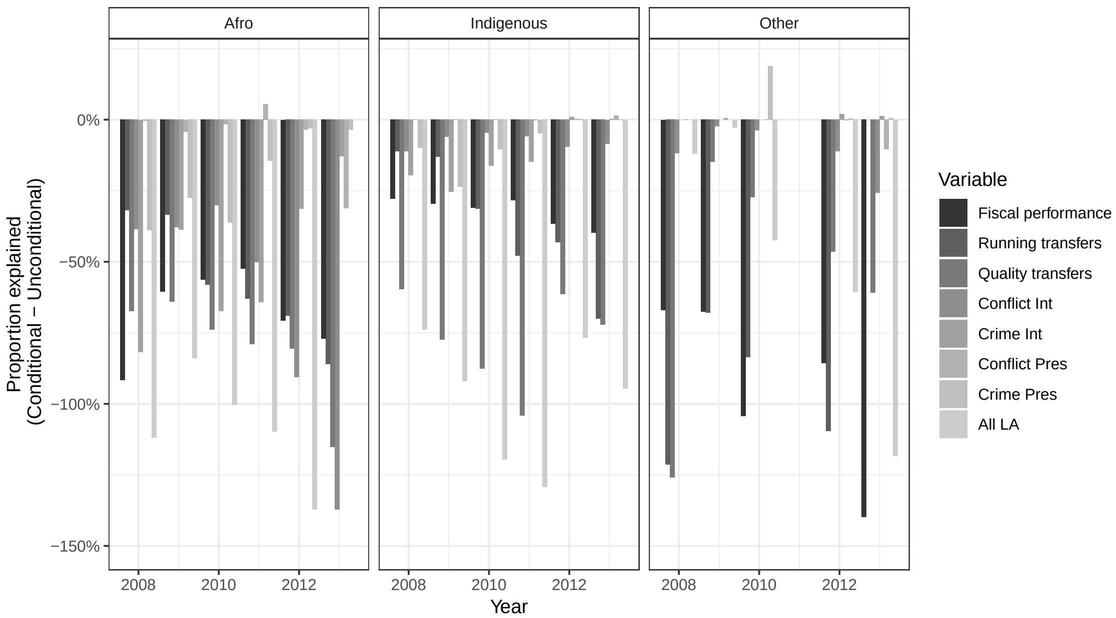

Figure A.38: Estimated changes in the LA contextual effect of ethnicity by LA characteristic for the sample of schools that are observed during all years

A.7 Appendix to Chapter 9

A.7.1 Likelihood Ratio Tests for Model Construction

This appendix presents the p-values for the χ^2 tests that were used to decide whether or not to include interaction terms in the models estimated in chapter 9.

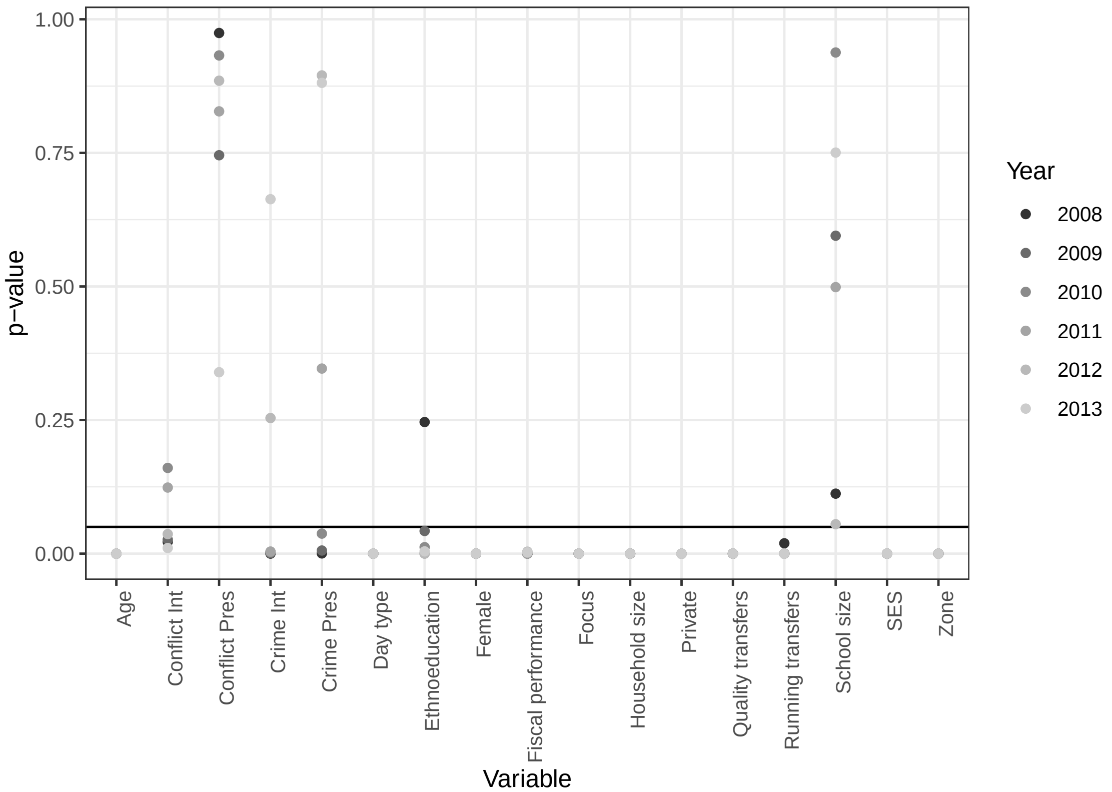

Figure A.39: P-values for the χ^2 test for competing student, school and LA variables predicting maths achievement for all years in the sample

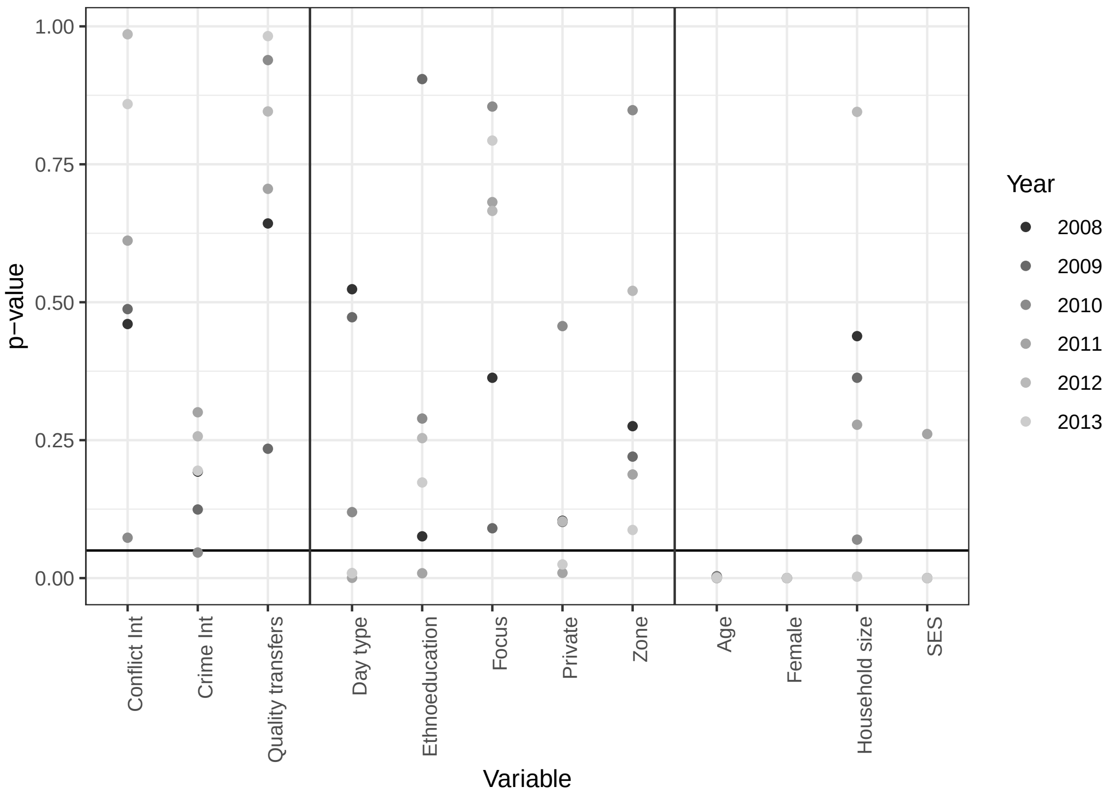

Figure A.40: P-values for the χ^2 test for the same-level interaction terms between student, school and LA and the within-school ethnic achievement gap and the school and LA contextual effect of all minority groups, respectively, for all years in the sample

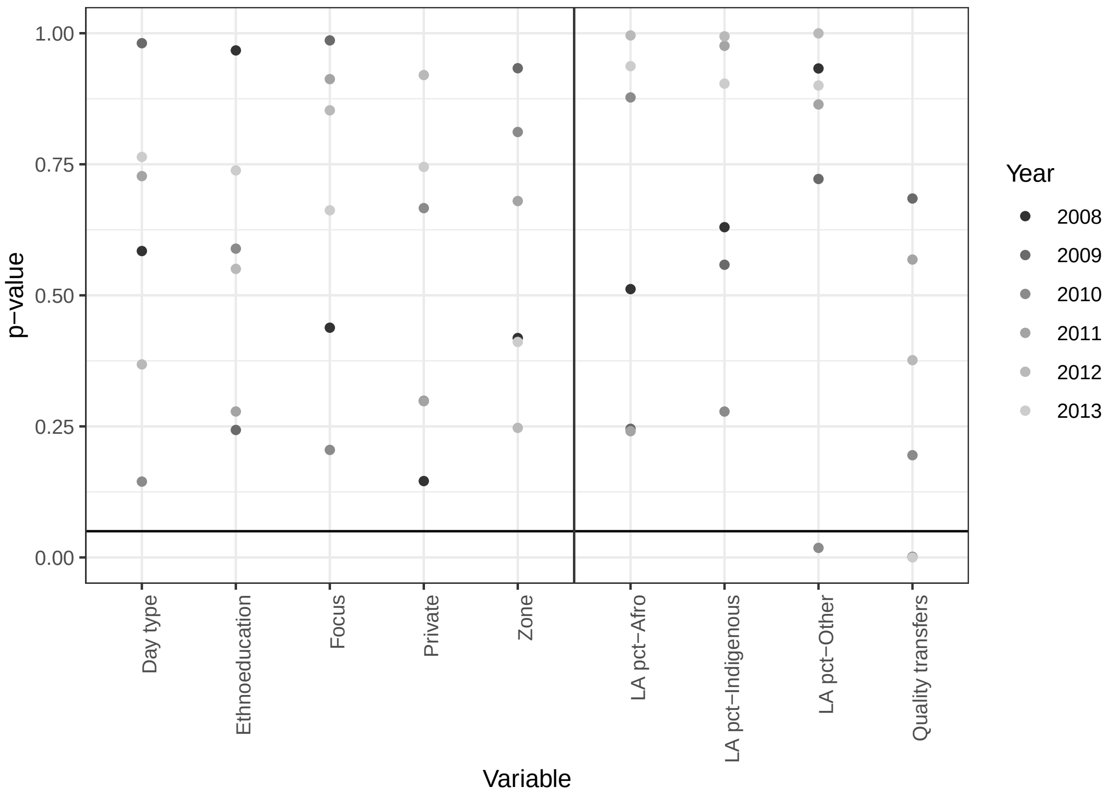

Figure A.41: P-values for the χ^2 test for the cross-level interaction terms between school and LA variables and the within-school ethnic achievement gap for all years in the sample

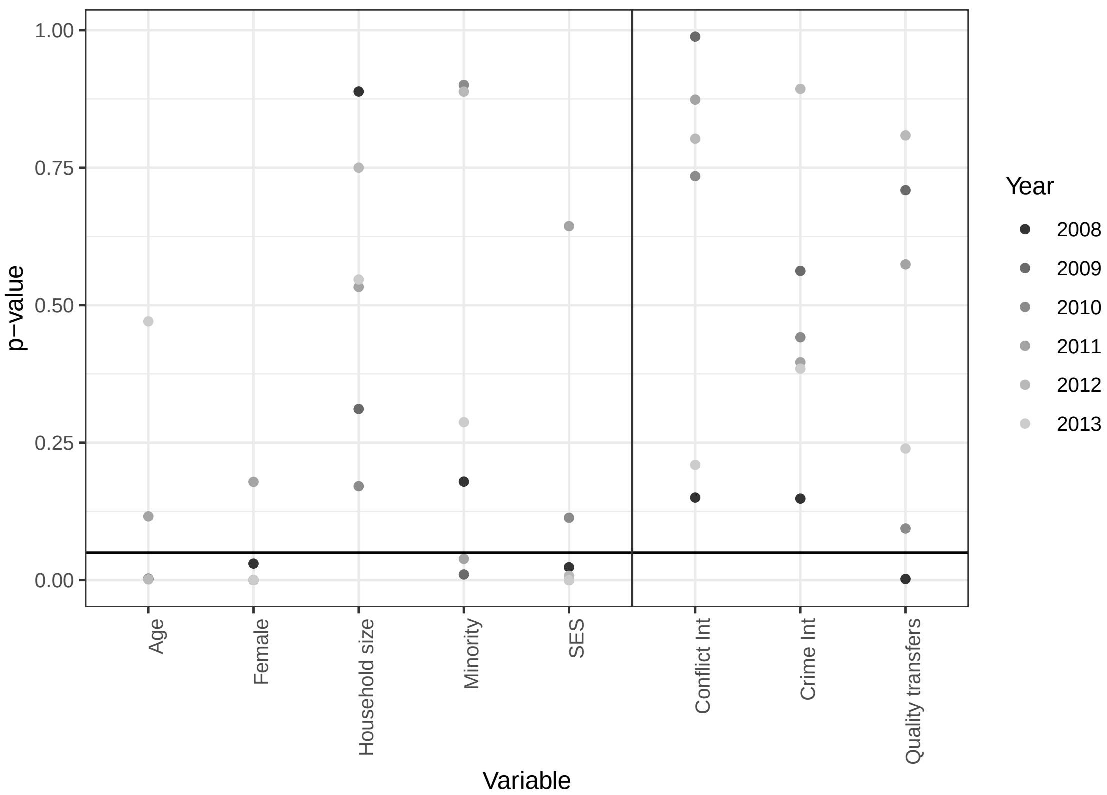

Figure A.42: P-values for the χ^2 test for the cross-level interaction terms between student and LA variables and the school contextual effect of ethnicity for all years in the sample

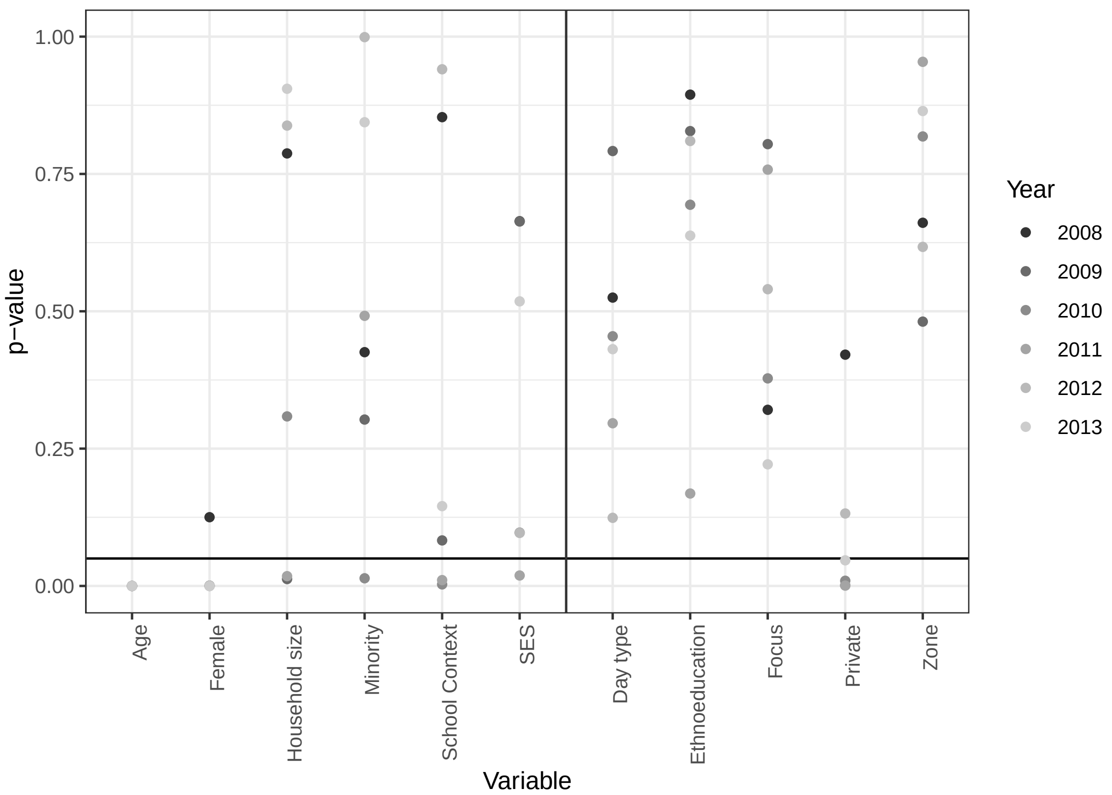

Figure A.43: P-values for the χ^2 test for the cross-level interaction terms between student and school variables and the LA contextual effect of ethnicity for all years in the sample

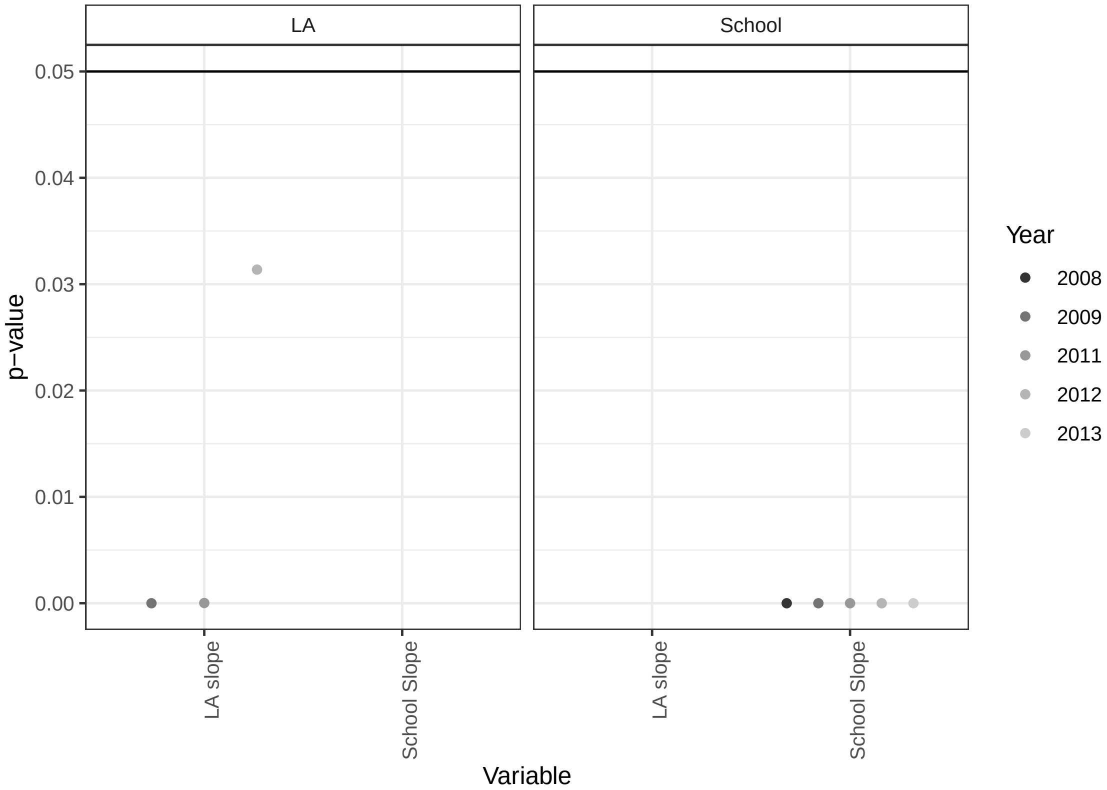

Figure A.44: P-values for the χ^2 test for the inclusion of random slopes for all years in the sample. Models that did not converge have been omitted from the figure

A.7.2 Estimation Results for the Models in Chapter 9

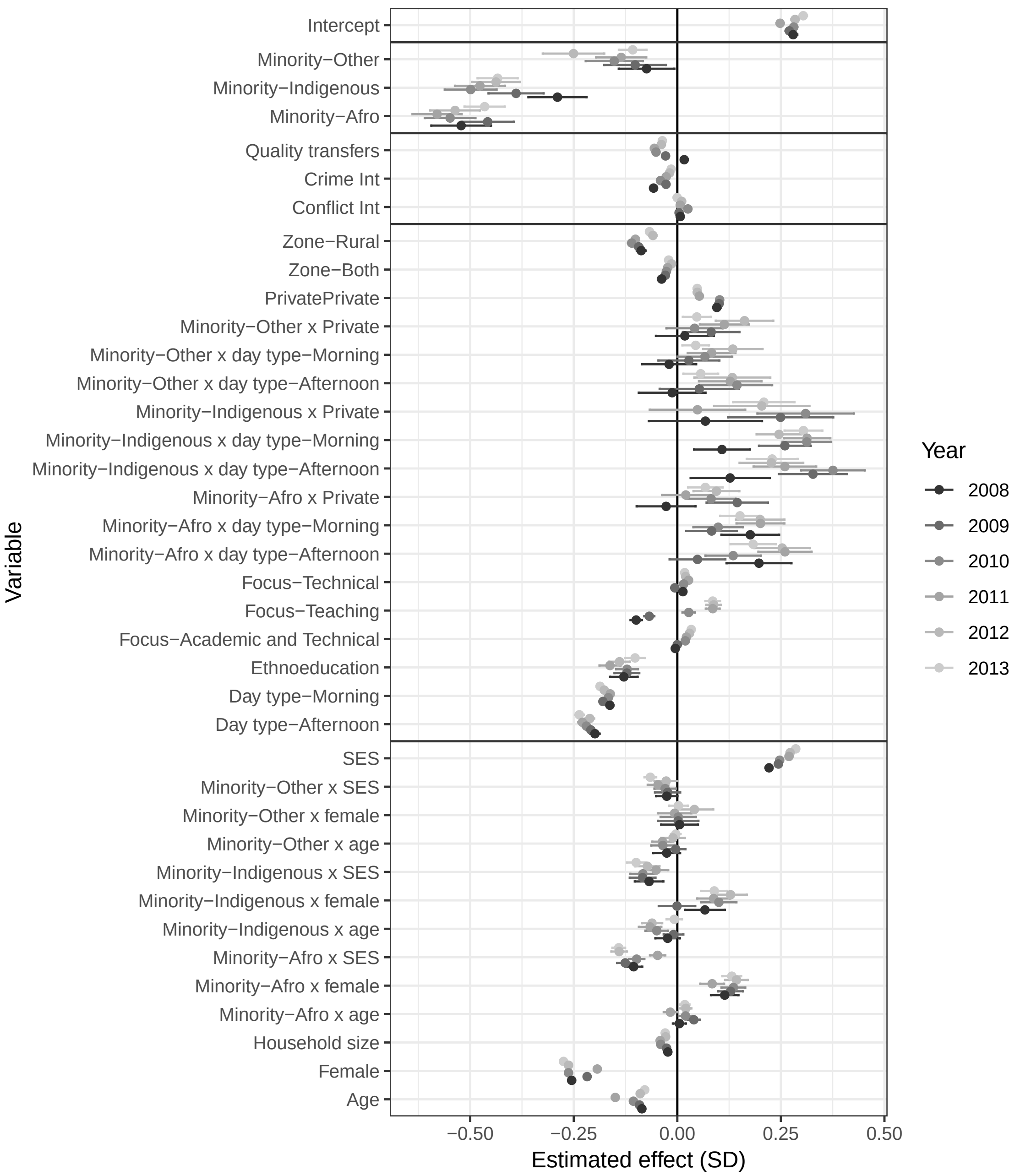

Figure A.45: Estimation results for model (9.2)

Table A.18: Estimated contextual effect multilevel models in chapter 9 for 2008

	(A.15)	(9.4)	(9.5)
Intercept	0.293***	0.299***	0.302***
	(0.021)	(0.023)	(0.022)
White			
Afro	-0.095***	-0.112***	-0.112***
	(0.016)	(0.024)	(0.025)
Indigenous	-0.051***	-0.083***	-0.056*
	(0.018)	(0.028)	(0.029)
Other	-0.046***	-0.055***	-0.049**
	(0.014)	(0.02)	(0.021)
School pct-Afro	-0.171***	-0.511***	-0.521***
	(0.039)	(0.119)	(0.118)
School pct-Indigenous	-0.126***	-0.189	-0.219**
	(0.046)	(0.126)	(0.101)
School pct-Other	-0.044	0.06	0.036
	(0.046)	(0.104)	(0.1)
LA pct-Afro	-0.115	-0.263*	-0.268*
	(0.102)	(0.14)	(0.139)
LA pct-Indigenous	-0.176	0	0.218
	(0.223)	(0.386)	(0.365)
LA pct-Other	-0.212	-0.103	-0.315
	(0.299)	(0.514)	(0.482)
Female	-0.273***	-0.284***	-0.284***
	(0.003)	(0.004)	(0.004)
SES	0.116***	0.118***	0.118***
	(0.002)	(0.002)	(0.002)
Household size	-0.015***	-0.014***	-0.014***
	(0.002)	(0.002)	(0.002)
Age	-0.076***	-0.079***	-0.079***

Continued on next page

Table A.18 – *Continued*

	(A.15)	(9.4)	(9.5)
	(0.002)	(0.002)	(0.002)
State			
Private	0.166***	0.166***	0.167***
	(0.012)	(0.016)	(0.015)
Urban			
Rural	-0.099***	-0.098***	-0.099***
	(0.013)	(0.013)	(0.013)
Both	-0.058***	-0.057***	-0.05***
	(0.016)	(0.016)	(0.015)
Academic			
Academic and Technical	-0.027**	-0.027**	-0.028**
	(0.013)	(0.013)	(0.013)
Teaching	-0.115***	-0.112***	-0.117***
	(0.02)	(0.02)	(0.02)
Technical	-0.036***	-0.034***	-0.035***
	(0.012)	(0.012)	(0.012)
Ethnoeducation	-0.051	-0.046	-0.039
	(0.042)	(0.044)	(0.04)
Crime Int	-0.049***	-0.048***	-0.048***
	(0.014)	(0.015)	(0.015)
Conflict Int	0.018	0.017	0.017
	(0.012)	(0.012)	(0.012)
Quality transfers	0.003	0	-0.001
	(0.019)	(0.02)	(0.02)
Afro x female		0.032	0.033
		(0.033)	(0.033)
Indigenous x female		0.057	0.062*
		(0.037)	(0.038)
Other x female		0.021	0.025
		(0.028)	(0.029)

Continued on next page

Table A.18 – *Continued*

	(A.15)	(9.4)	(9.5)
Afro x SES		-0.029	-0.028
		(0.018)	(0.018)
Indigenous x SES		-0.028	-0.011
		(0.023)	(0.023)
Other x SES		-0.032**	-0.025*
		(0.014)	(0.014)
Afro x age		-0.014	-0.014
		(0.013)	(0.013)
Indigenous x age		-0.017	-0.027
		(0.019)	(0.02)
Other x age		-0.016	-0.017
		(0.016)	(0.016)
School pct-Afro x day type-Morning		0.294***	0.291***
		(0.105)	(0.104)
School pct-Afro x day type-Afternoon		0.256**	0.262**
		(0.124)	(0.123)
School pct-Indigenous x day type-Morning		0.073	0.079
		(0.103)	(0.077)
School pct-Indigenous x day type-Afternoon		0.099	0.09
		(0.157)	(0.113)
School pct-Other x day type-Morning		-0.094	-0.083
		(0.104)	(0.1)
School pct-Other x day type-Afternoon		-0.049	-0.045
		(0.147)	(0.14)
School pct-Afro x age		0.051***	0.051***
		(0.018)	(0.018)
School pct-Afro x female		0.067	0.067
		(0.043)	(0.043)
School pct-Afro x SES		-0.041*	-0.041*
		(0.023)	(0.023)

Continued on next page

Table A.18 – *Continued*

	(A.15)	(9.4)	(9.5)
School pct-Indigenous x age		0.035	0.048
		(0.03)	(0.03)
School pct-Indigenous x female		0.021	0.019
		(0.051)	(0.052)
School pct-Indigenous x SES		0.027	0.014
		(0.033)	(0.034)
School pct-Other x age		-0.059**	-0.057**
		(0.029)	(0.029)
School pct-Other x female		-0.033	-0.038
		(0.047)	(0.047)
School pct-Other x SES		0.019	0.013
		(0.027)	(0.027)
LA pct-Afro x age		-0.006	-0.006
		(0.02)	(0.02)
LA pct-Afro x female		0.086**	0.086**
		(0.043)	(0.043)
LA pct-Afro x Private		0.166	0.17
		(0.136)	(0.136)
LA pct-Afro x School pct-Afro		0.117	0.137
		(0.15)	(0.149)
LA pct-Afro x School pct-Indigenous		0.187	0.192
		(0.381)	(0.306)
LA pct-Afro x School pct-Other		0.086	0.117
		(0.421)	(0.403)
LA pct-Indigenous x age		-0.412***	-0.411***
		(0.054)	(0.054)
LA pct-Indigenous x female		-0.003	0.002
		(0.089)	(0.089)
LA pct-Indigenous x Private		0.515	0.449
		(0.41)	(0.399)

Continued on next page

	(A.15)	(9.4)	(9.5)
LA pct-Indigenous x School pct-Afro		-0.222	-0.197
		(0.61)	(0.606)
LA pct-Indigenous x School pct-Indigenous		-0.172	-0.392
		(0.525)	(0.459)
LA pct-Indigenous x School pct-Other		-0.135	-0.061
		(1.181)	(1.106)
LA pct-Other x age		0.423***	0.413***
		(0.066)	(0.066)
LA pct-Other x female		0.213	0.214
		(0.147)	(0.147)
LA pct-Other x Private		-0.628	-0.615
		(0.617)	(0.59)
LA pct-Other x School pct-Afro		-0.391	-0.224
		(1.556)	(1.542)
LA pct-Other x School pct-Indigenous		-0.021	0.181
		(0.901)	(0.744)
LA pct-Other x School pct-Other		-0.619	-0.47
		(0.997)	(0.919)
Day type-Morning	-0.189***	-0.194***	-0.194***
	(0.012)	(0.012)	(0.012)
Day type-Afternoon	-0.215***	-0.221***	-0.221***
	(0.017)	(0.017)	(0.018)
σ^2_v	0.012	0.012	0.012
	(0.002)	(0.002)	(0.002)
$\sigma^2_{u_O}$			0.007
			(0.006)
$\sigma^2_{u_1}$	0.101	0.101	0.103
	(0.002)	(0.002)	(0.002)
$\sigma^2_{u_I}$			0.045
			(0.011)

Continued on next page

Table A.18 – *Continued*

	(A.15)	(9.4)	(9.5)
$\sigma^2_{u_A}$			0
			(0)
$\sigma_{u_{O1}}$			-0.01
			(0.005)
$\sigma_{u_{I1}}$			-0.058
			(0.006)
$\sigma_{u_{A1}}$			0
			(0)
$\sigma_{u_{OI}}$			-0.005
			(0.012)
$\sigma_{u_{OA}}$			0
			(0)
$\sigma_{u_{IA}}$			0
			(0)
σ^2_r	0.785	0.785	0.785
	(0.002)	(0.002)	(0.002)
VPC_{LA}	1.3	1.32	
$VPC_{schools}$	11.27	11.23	
$R^2_{v_0}$	72.48	72.11	
$R^2_{u_0}$	42.17	42.47	
R^2_e	3.46	3.51	
R^2	12.87	12.95	
Deviance	1,019,621	1,019,401	1,019,335

Standard errors in parenthesis.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Dependent variable: maths.

Sample of 80 LAs, 47,010 Schools, 387,773 Students.

M_{ij}: Minority student, $\overline{M}_{.j}$: School proportion of minority students.

The first model in Table A.18 is a contextual effect model including all control variables but no interaction effects, given by

$$y_{ijk} = \alpha + \theta_A^W A_{ijk} + \theta_I^W I_{ijk} + \theta_O^W O_{ijk} + \theta^{CSA}\overline{A}_{.jk} + \theta^{CSI}\overline{I}_{.jk} + \theta^{CSO}\overline{O}_{.jk} + \theta^{CLA}\overline{A}_{..k} + \theta^{CLI}\overline{I}_{..k} + \theta^{CLO}\overline{O}_{..k} + \mathbf{x}_{ijk}^T\delta_{\mathbf{x}} + \mathbf{z}_{jk}^T\delta_{\mathbf{z}} + \mathbf{w}_k^T\delta_{\mathbf{w}} + v_k + u_{1_{jk}} + r_{ijk} \quad \text{(A.15)}$$

A.7.3 Estimaton Results for all Years in the Sample

This appendix presents the results of the estimated models for all years in the sample (2008 to 2013). Chapter 9 only presented the results for 2008.

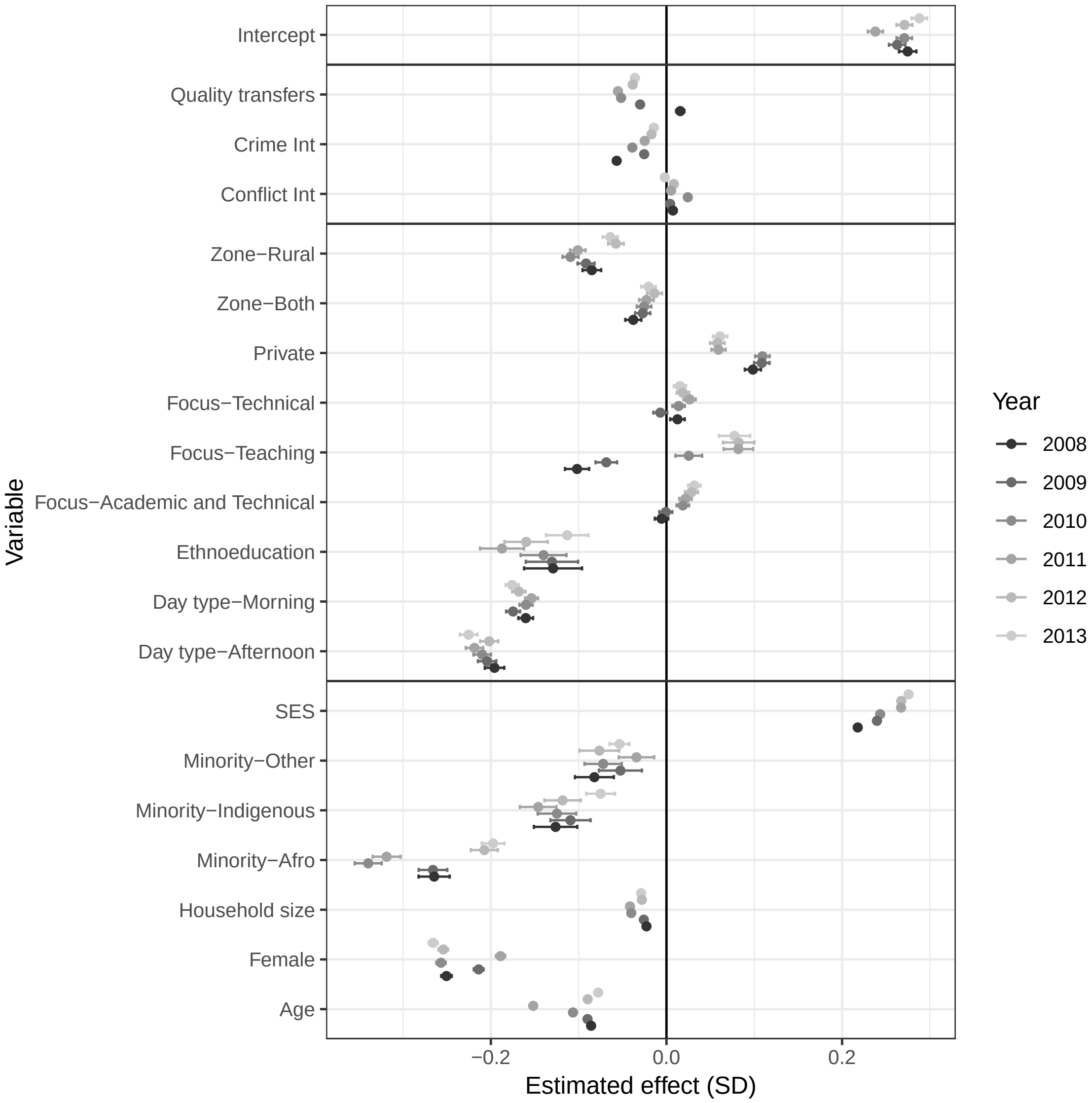

Figure A.46: Estimated parameters and 95% confidence intervals for the single-level linear effects model (9.1) for maths achievement by year

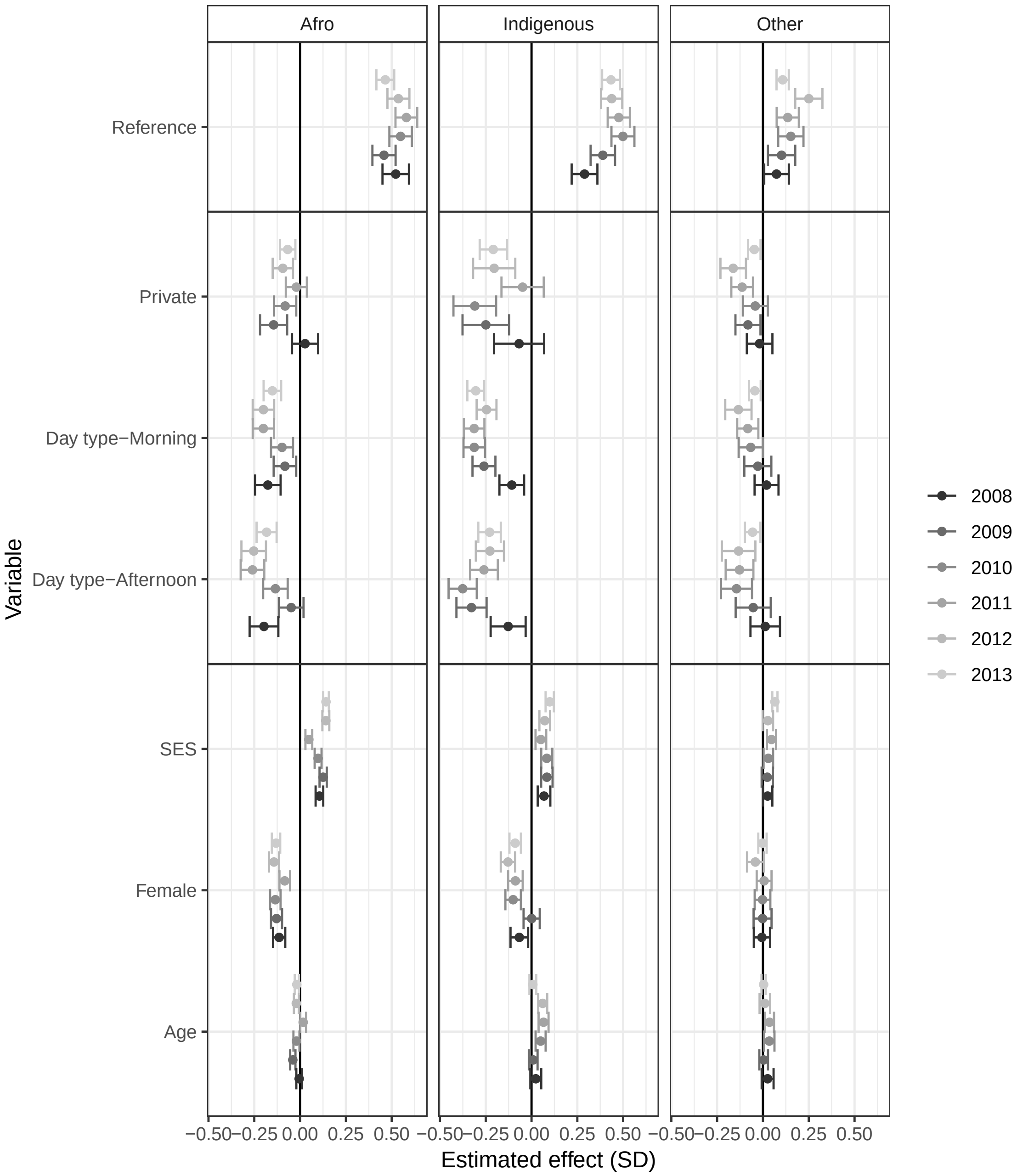

Figure A.47: Estimated interaction effects and 95% confidence intervals for the single-level model including interactions (9.2) for maths achievement by year

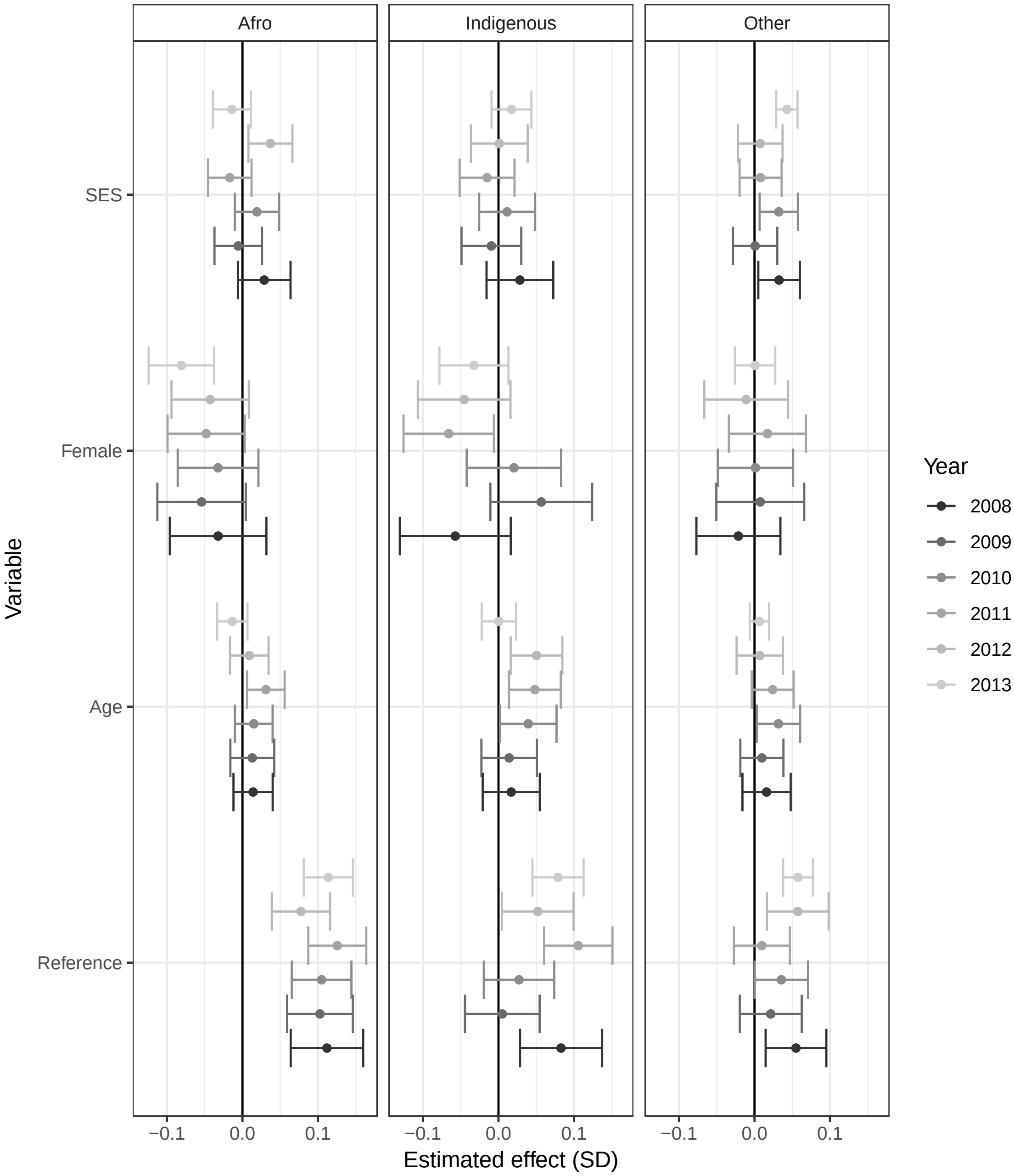

Figure A.48: Estimated interaction effects and 95% confidence intervals for the within-school gap according to the three-level random intercept model including interactions (9.4) for maths achievement by year

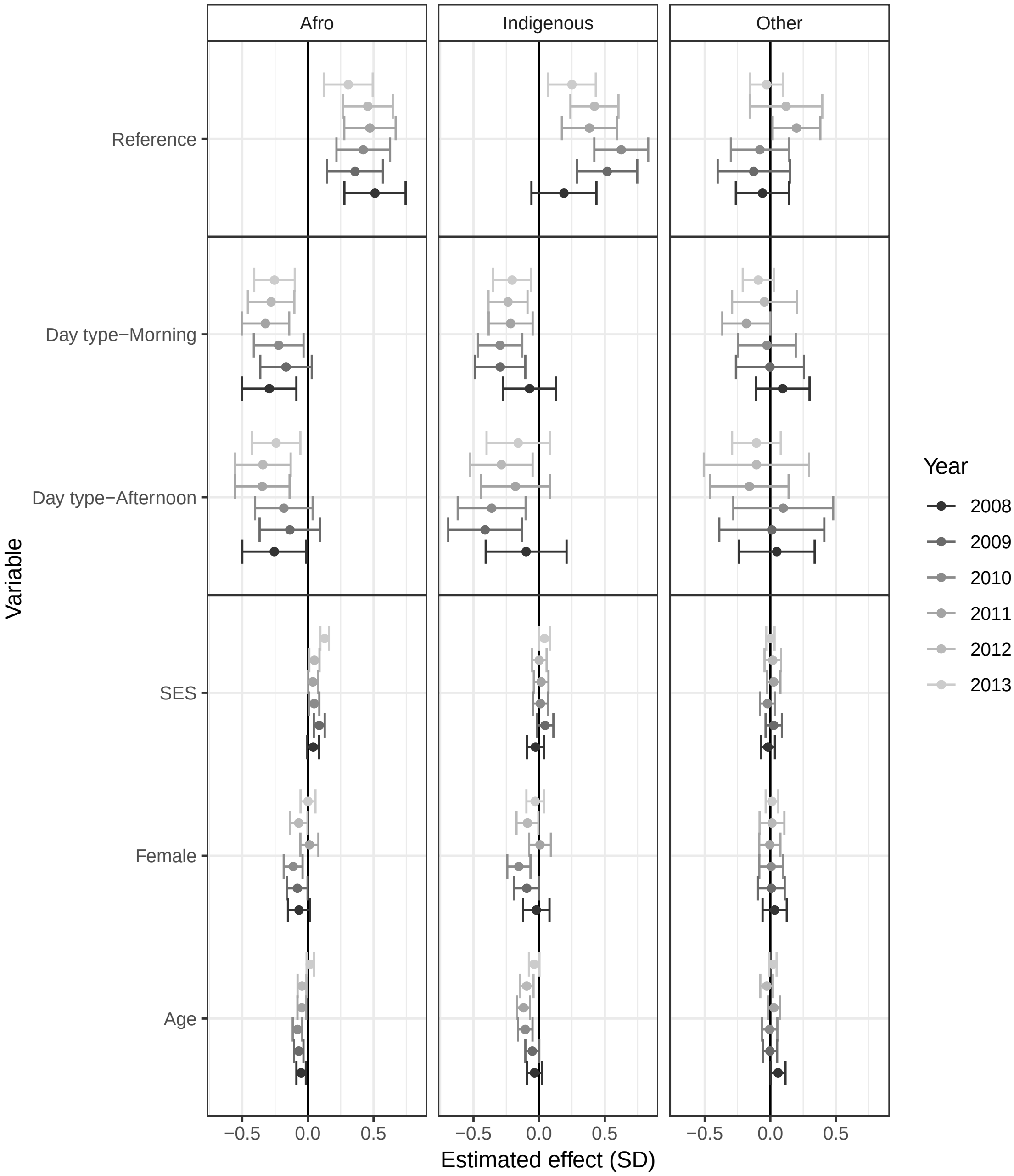

Figure A.49: Estimated interaction effects and 95% confidence intervals for the school contextual effect of ethnicity according to the three-level random intercept model including interactions (9.4) for maths achievement by year

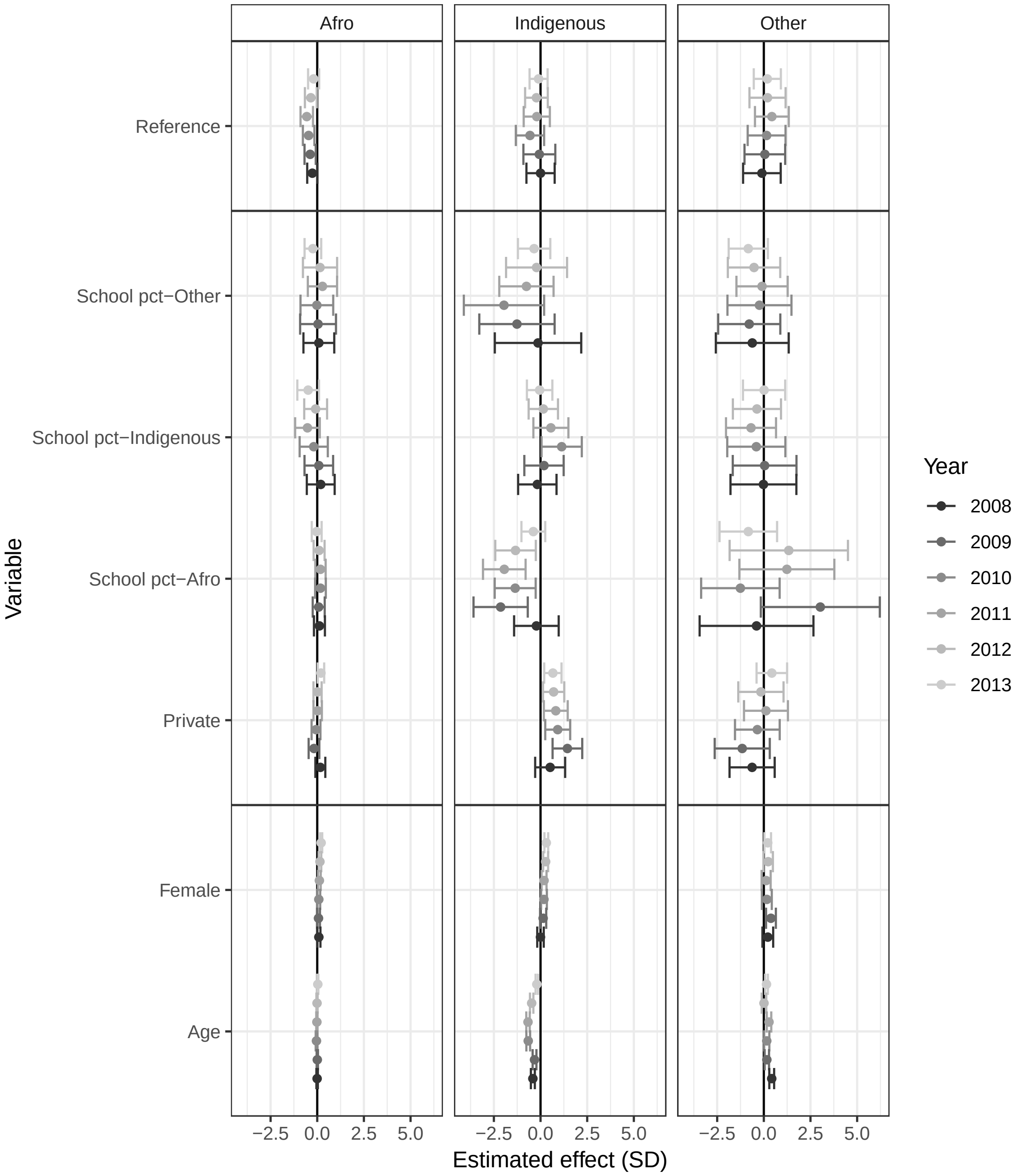

Figure A.50: Estimated interaction effects and 95% confidence intervals for the LA contextual effect of ethnicity according to the three-level random intercept model including interactions (9.4) for maths achievement by year

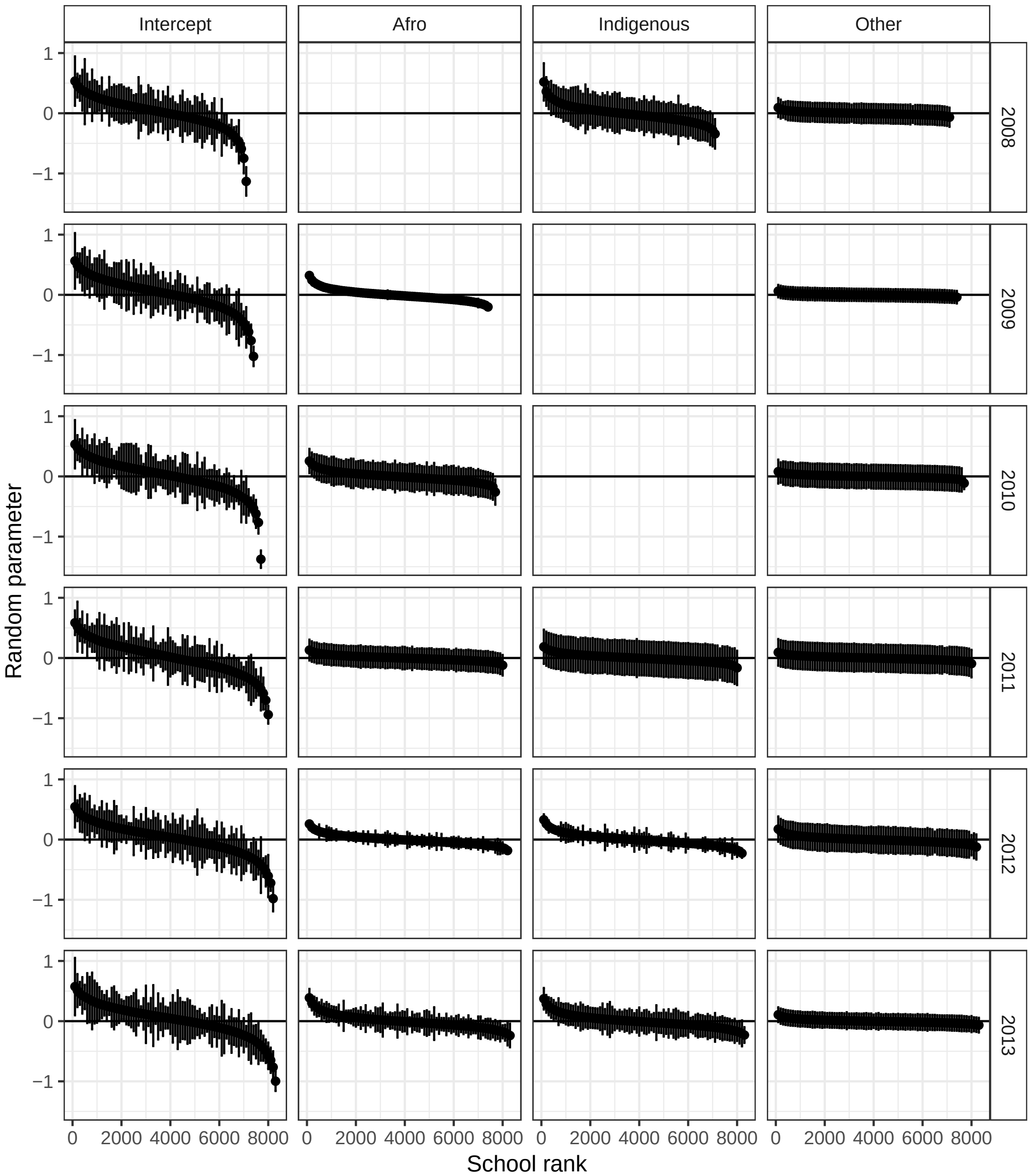

Figure A.51: Caterpillar plot for the random residual intercept and within-school ethnic achievement gap of the model given by equations (9.4) and (9.5) by year

A.7.4 Exploring the Assumption of Normality of the Residuals

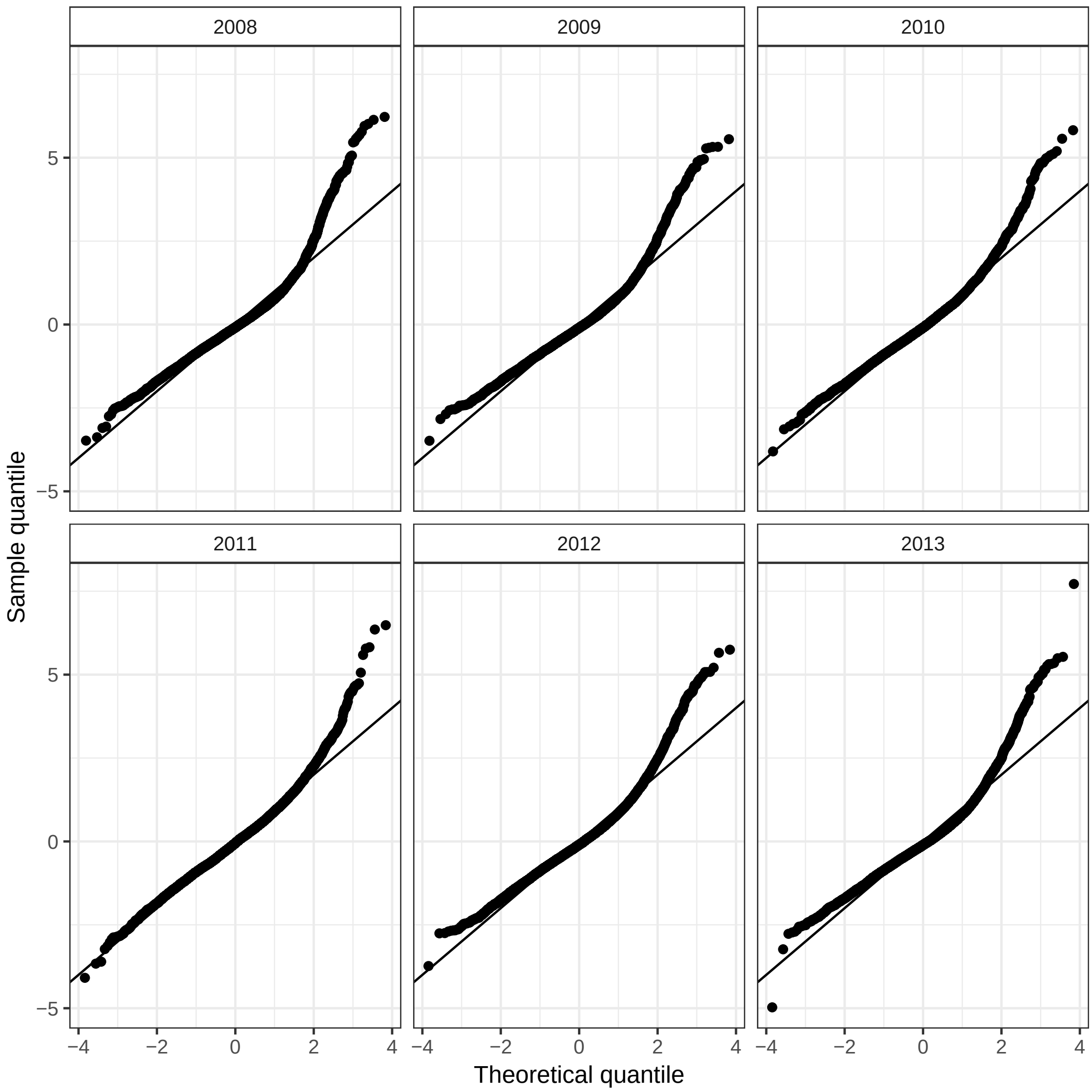

Figure A.52: Q-Q plot for the random residual intercept as estimated by the model given by equations (9.4) and (9.5) by year

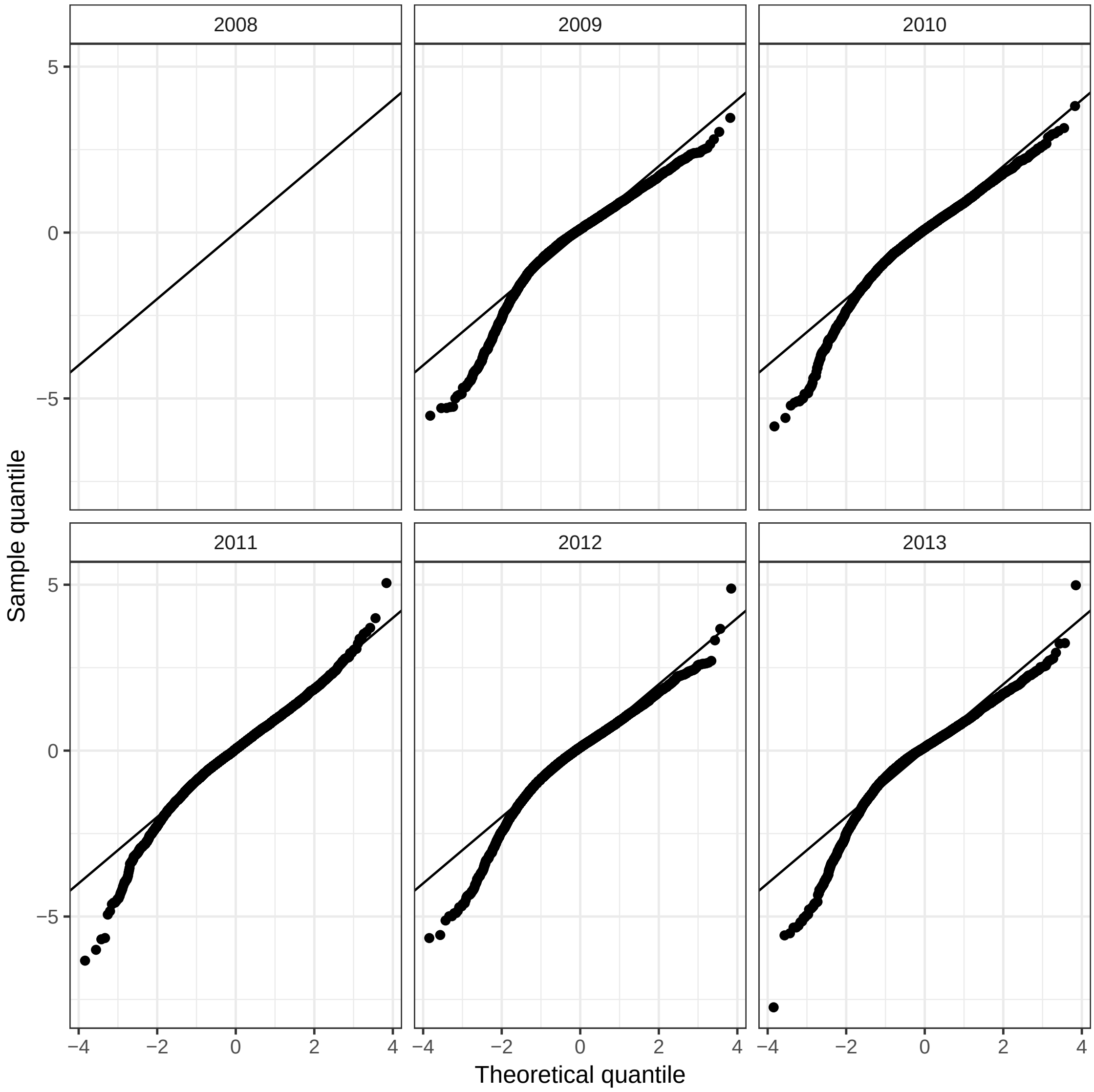

Figure A.53: Q-Q plot for the random residual random within-school gap between White and Afrocolombian students as estimated by the model given by equations (9.4) and (9.5) by year

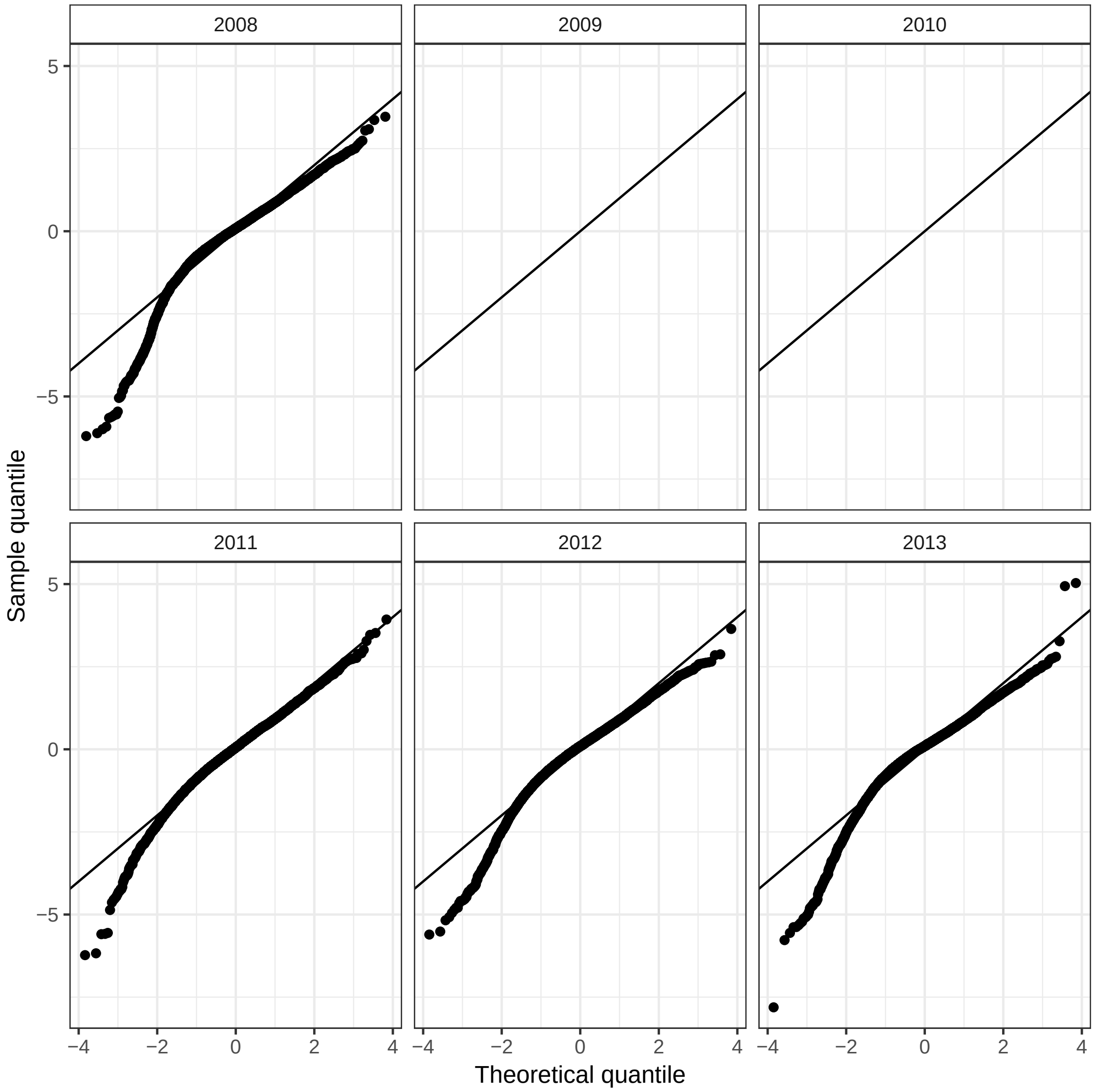

Figure A.54: Q-Q plot for the random residual random within-school gap between White and Indigenous students as estimated by the model given by equations (9.4) and (9.5) by year

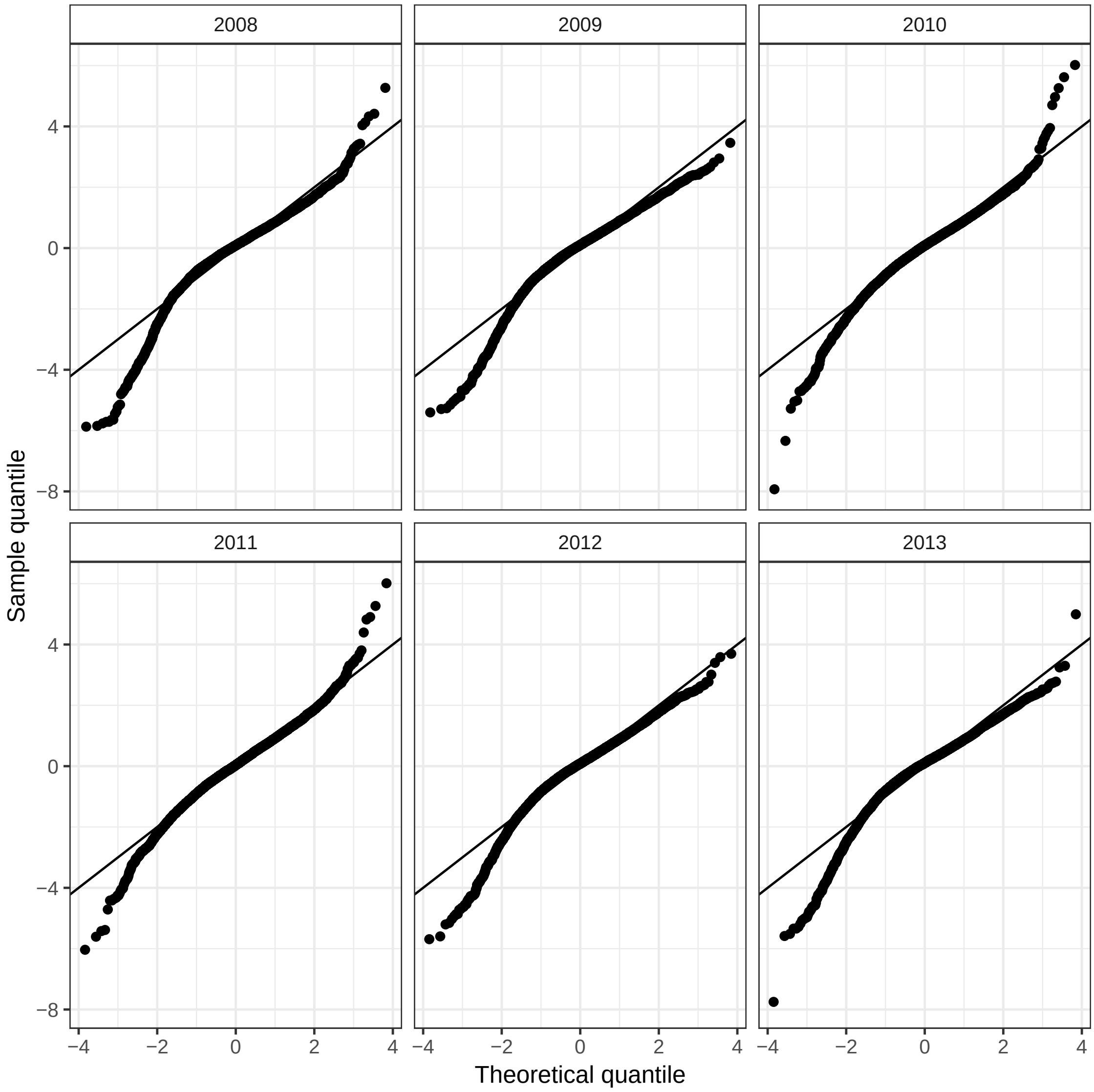

Figure A.55: Q-Q plot for the random residual random within-school gap between White and other minority students as estimated by the model given by equations (9.4) and (9.5) by year

A.7.5 Estimation Results for Ethnically Heterogeneous Schools

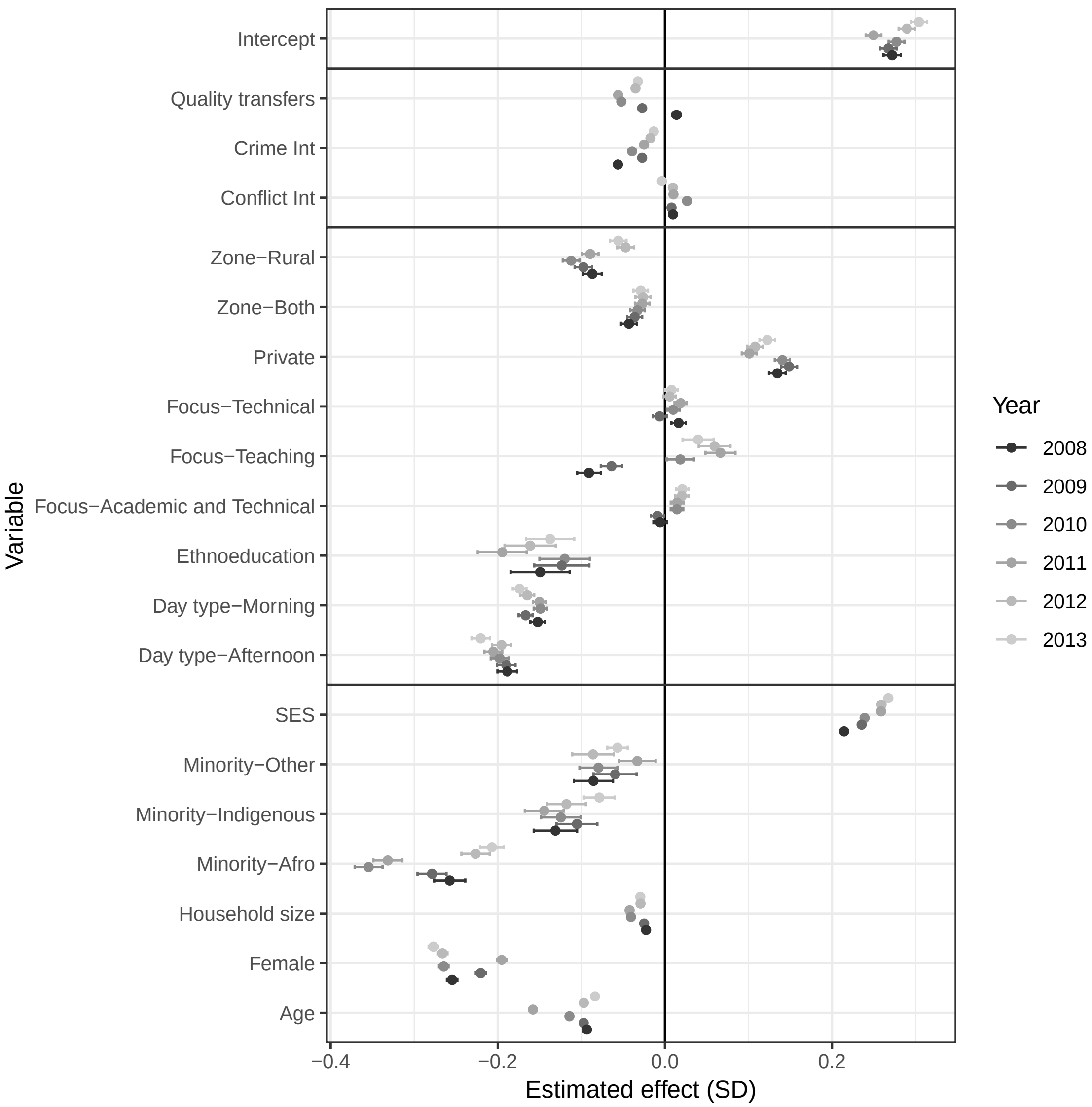

Figure A.56: Estimated parameters and 95% confidence intervals for the single-level linear effects model (9.1) for maths achievement by year for the sample of heterogeneous schools

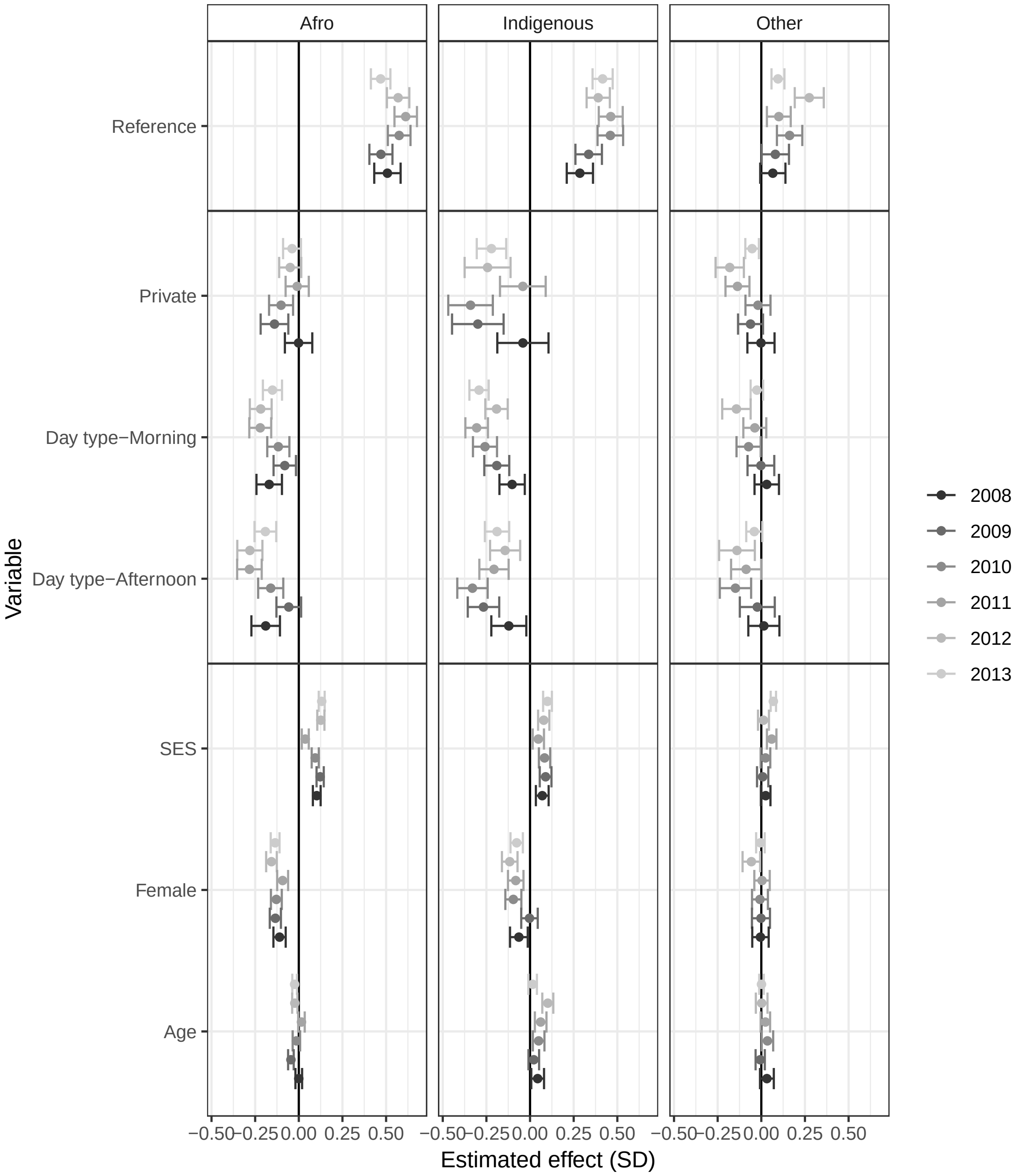

Figure A.57: Estimated interaction effects and 95% confidence intervals for the single-level model including interactions (9.2) for maths achievement by year for the sample of heterogeneous schools

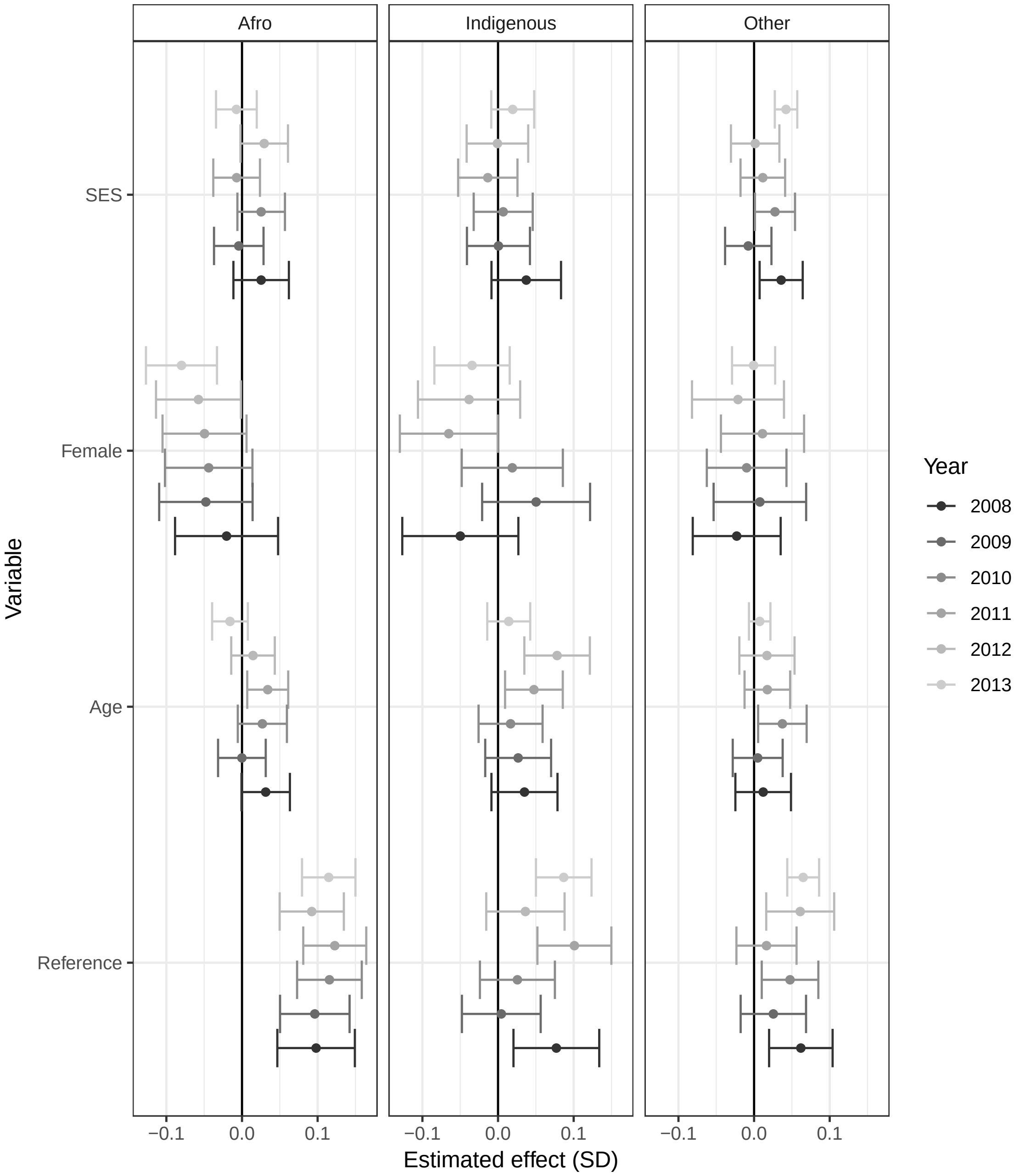

Figure A.58: Estimated interaction effects and 95% confidence intervals for the within-school gap according to the three-level random intercept model including interactions (9.4) for maths achievement by year for the sample of heterogeneous schools

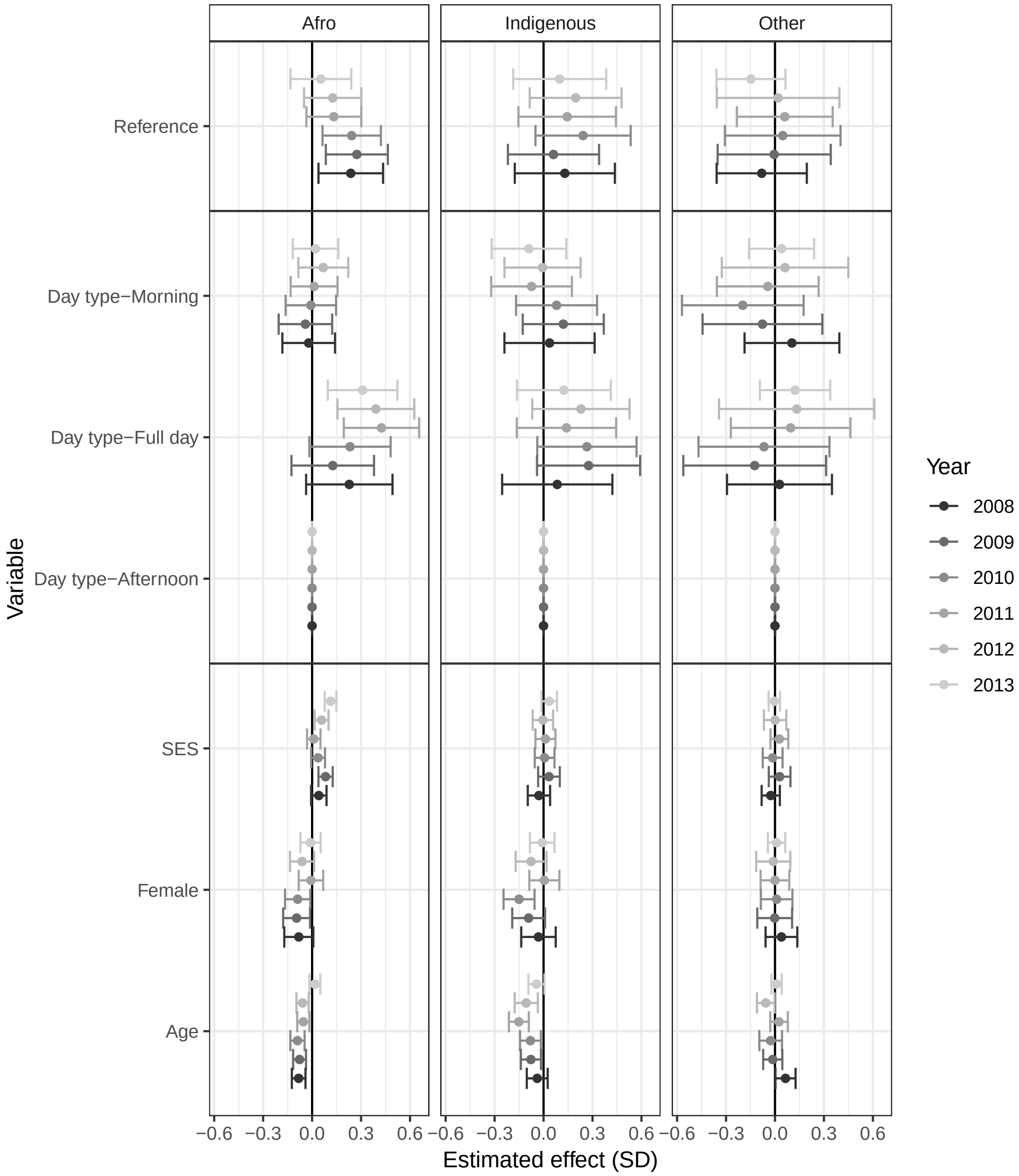

Figure A.59: Estimated interaction effects and 95% confidence intervals for the school contextual effect of ethnicity according to the three-level random intercept model including interactions (9.4) for maths achievement by year for the sample of heterogeneous schools

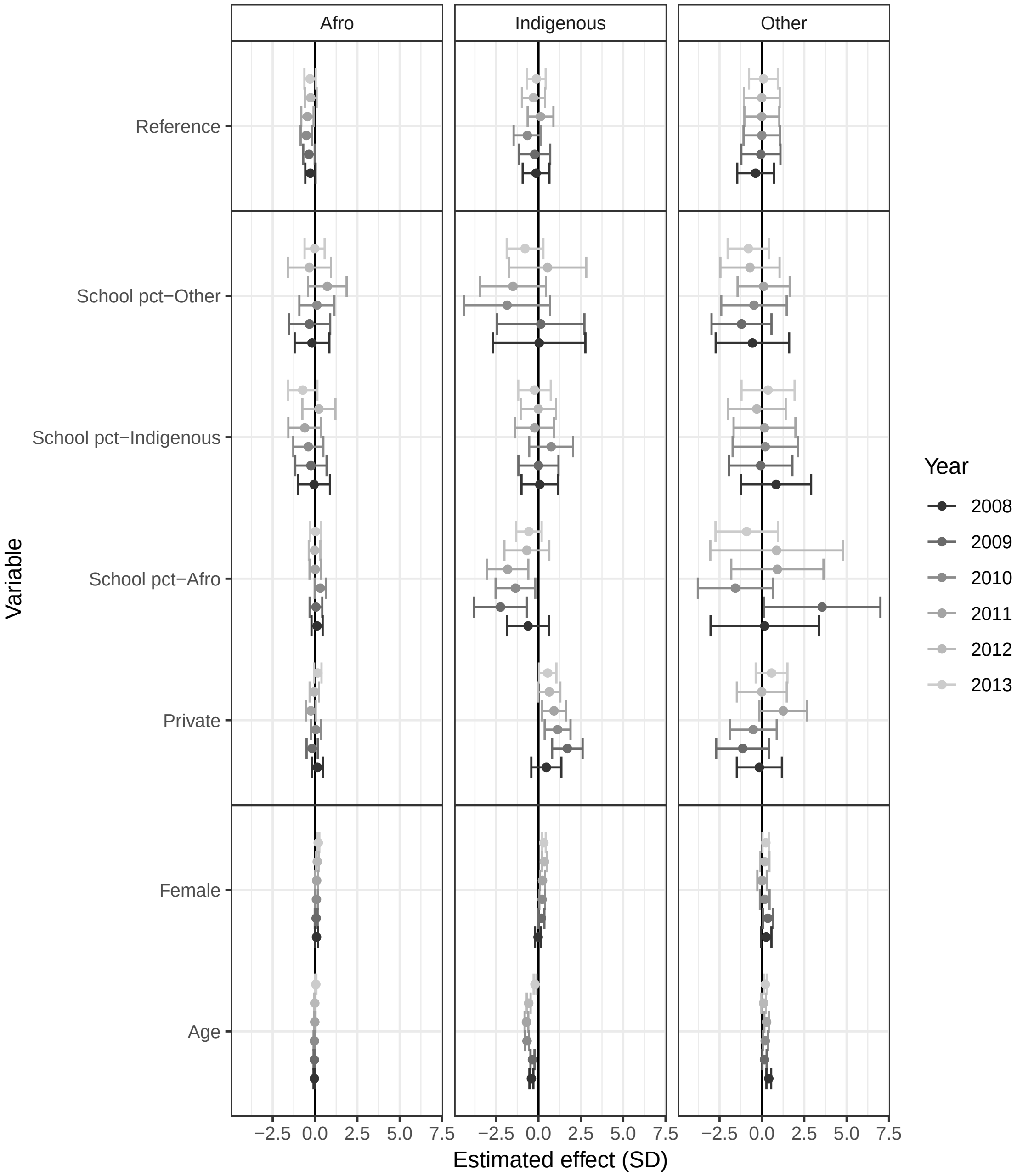

Figure A.60: Estimated interaction effects and 95% confidence intervals for the LA contextual effect of ethnicity according to the three-level random intercept model including interactions (9.4) for maths achievement by year for the sample of heterogeneous schools

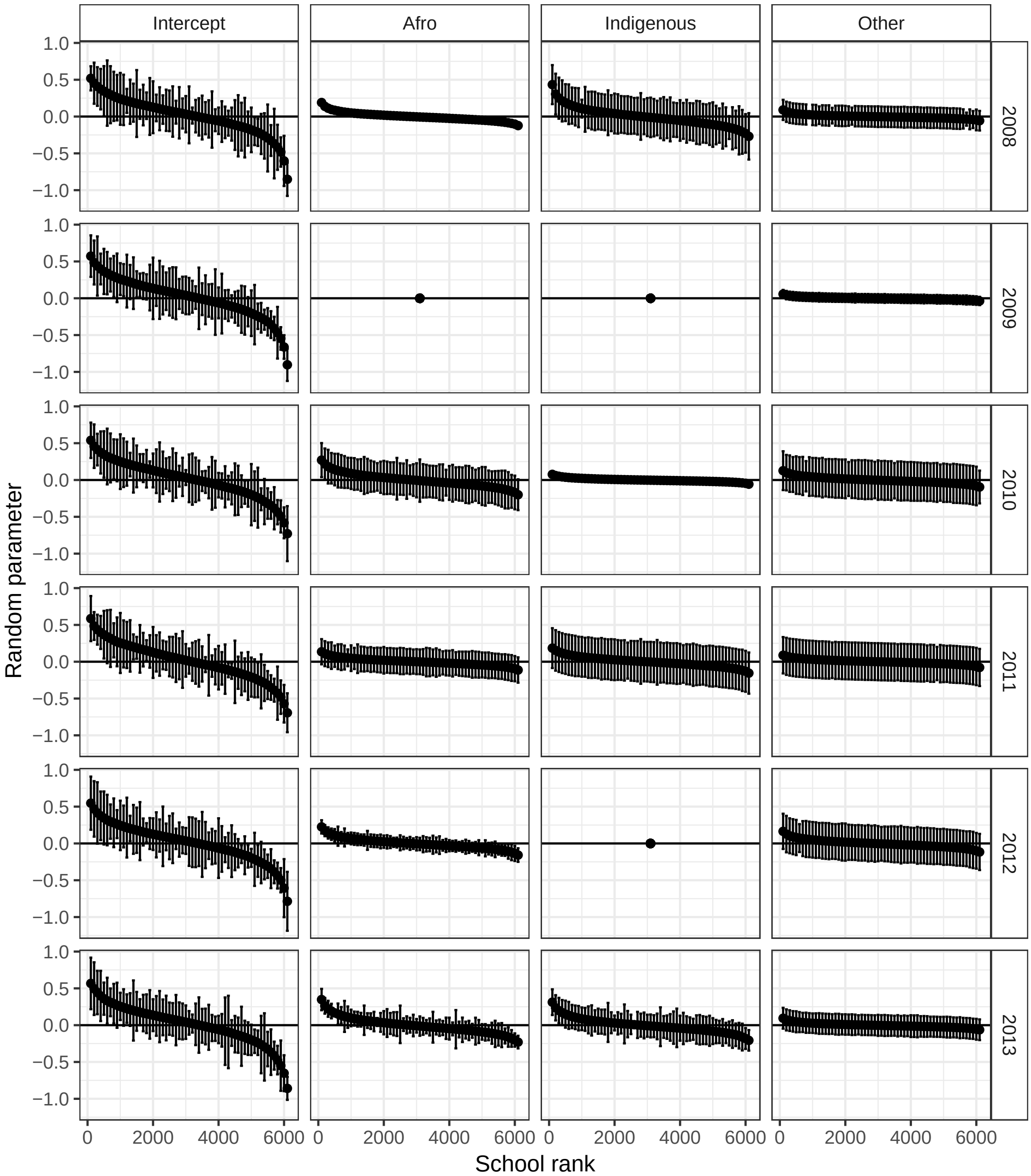

Figure A.61: Caterpillar plot for the random residual intercept and within-school ethnic achievement gap of the model given by equations (9.4) and (9.5) by year for the sample of heterogeneous schools

A.7.6 Estimation Results for Schools that are Observed During all Years in the Sample

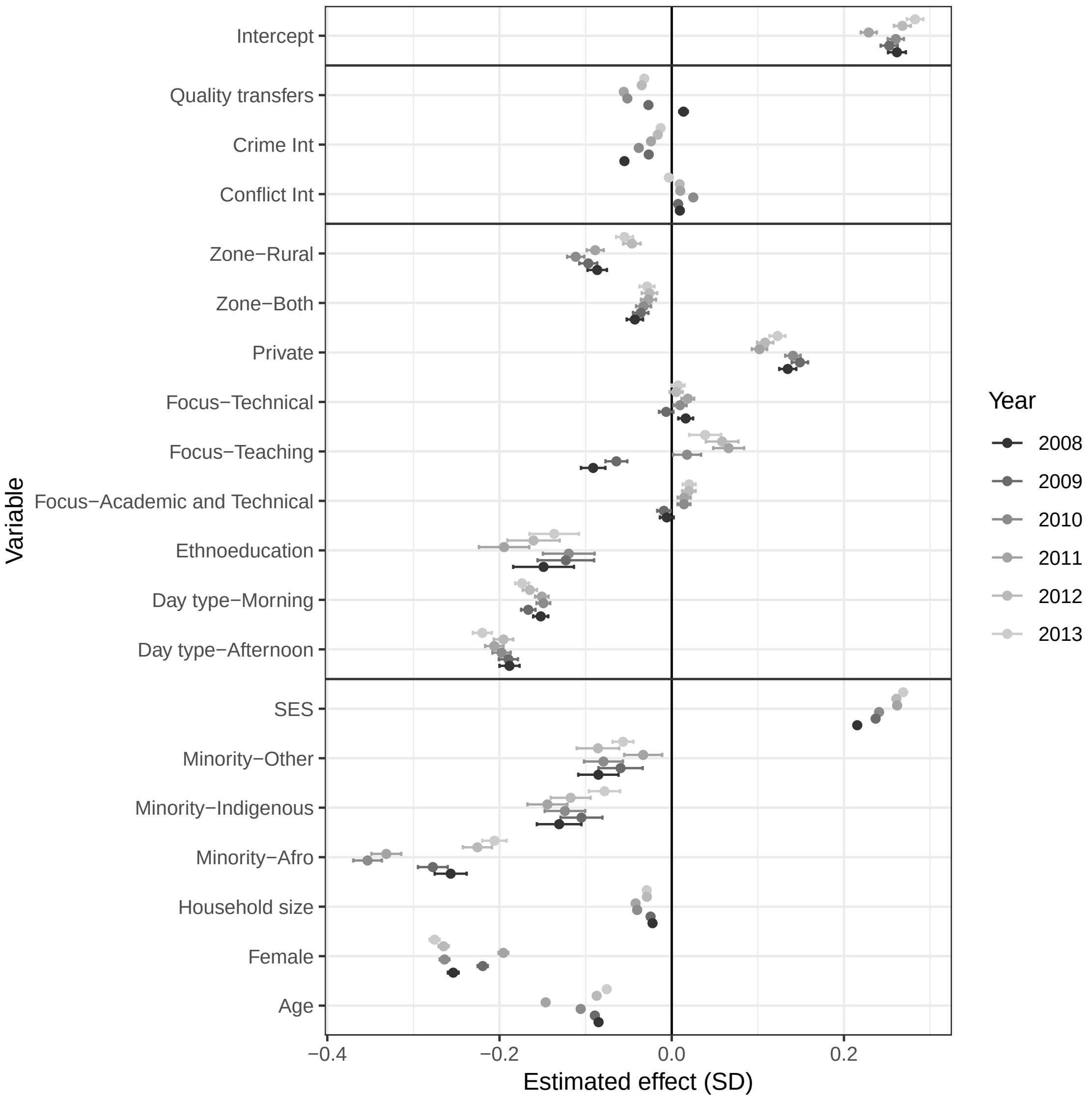

Figure A.62: Estimated parameters and 95% confidence intervals for the single-level linear effects model (9.1) for maths achievement by year for the sample of schools observed during all years

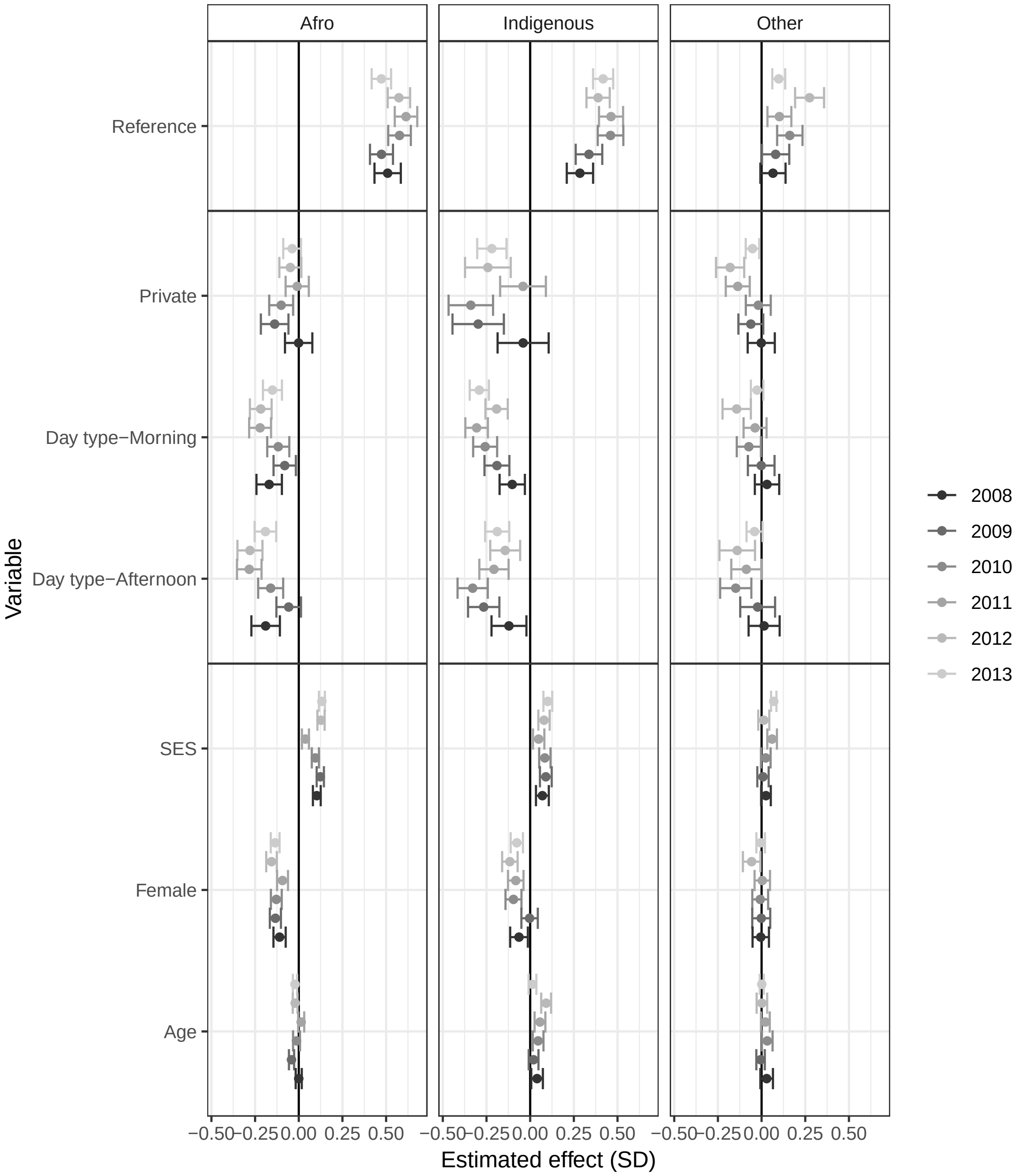

Figure A.63: Estimated interaction effects and 95% confidence intervals for the single-level model including interactions (9.2) for maths achievement by year for the sample of schools observed during all years

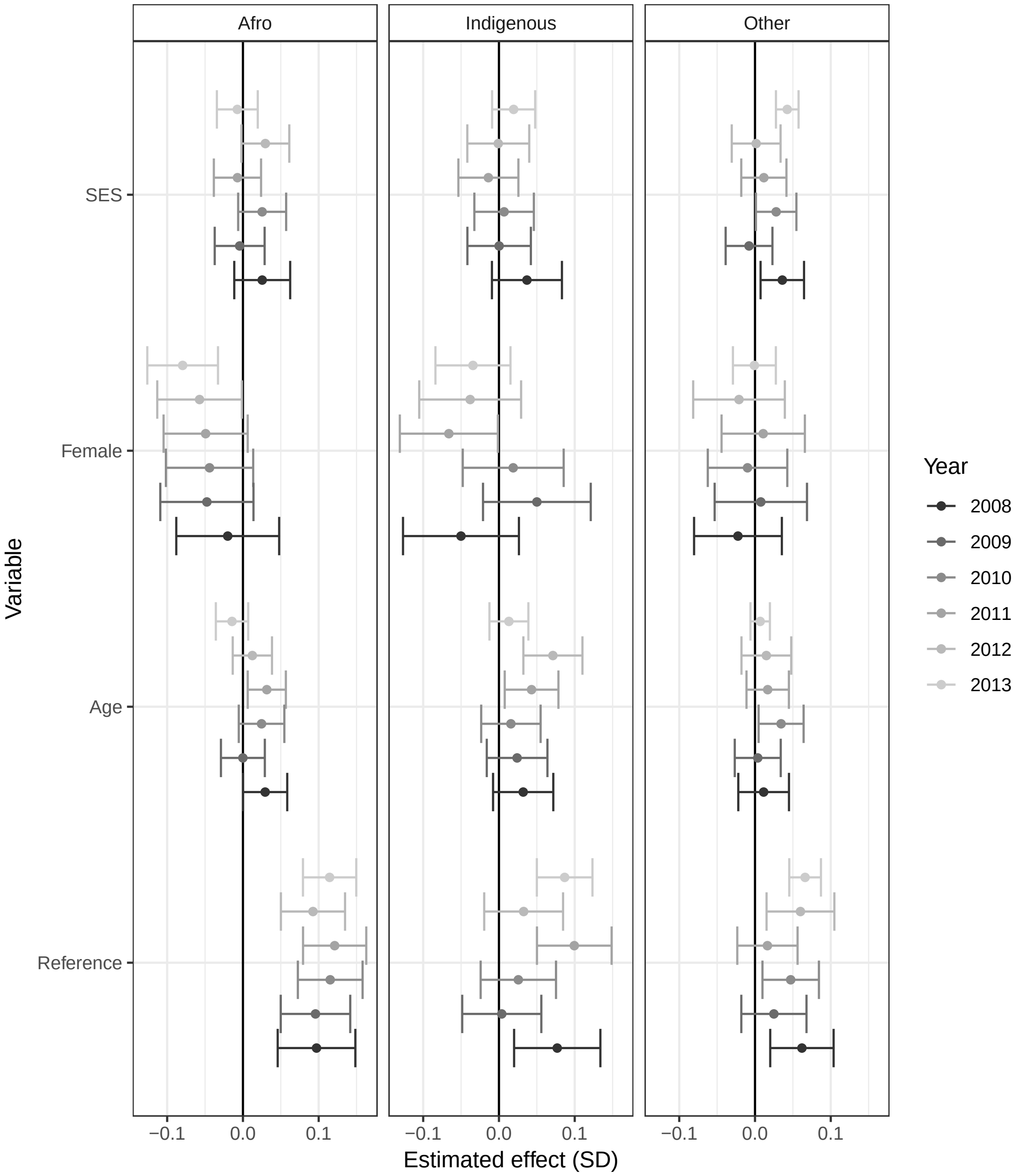

Figure A.64: Estimated interaction effects and 95% confidence intervals for the within-school gap according to the three-level random intercept model including interactions (9.4) for maths achievement by year for the sample of schools observed during all years

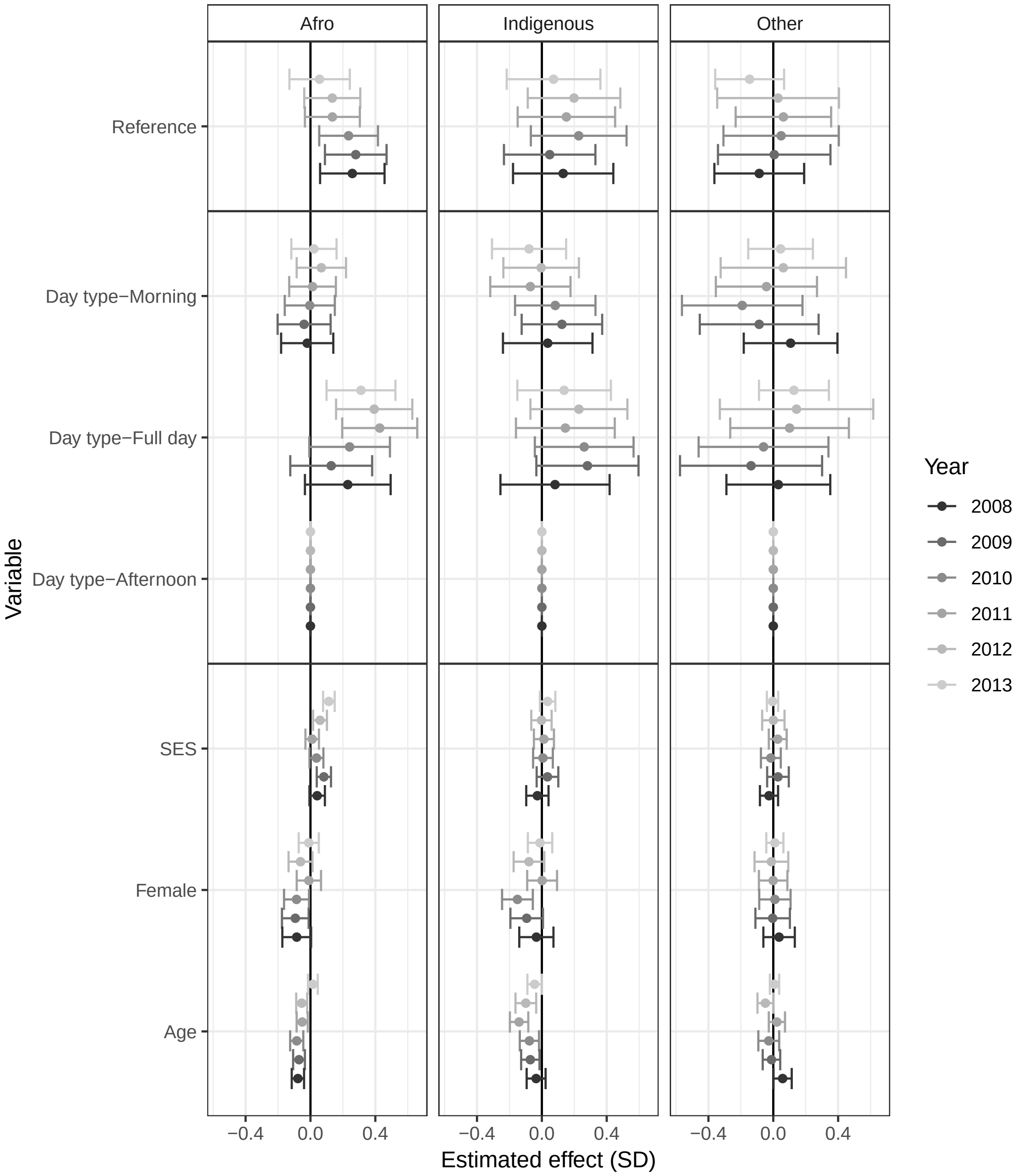

Figure A.65: Estimated interaction effects and 95% confidence intervals for the school contextual effect of ethnicity according to the three-level random intercept model including interactions (9.4) for maths achievement by year for the sample of schools observed during all years

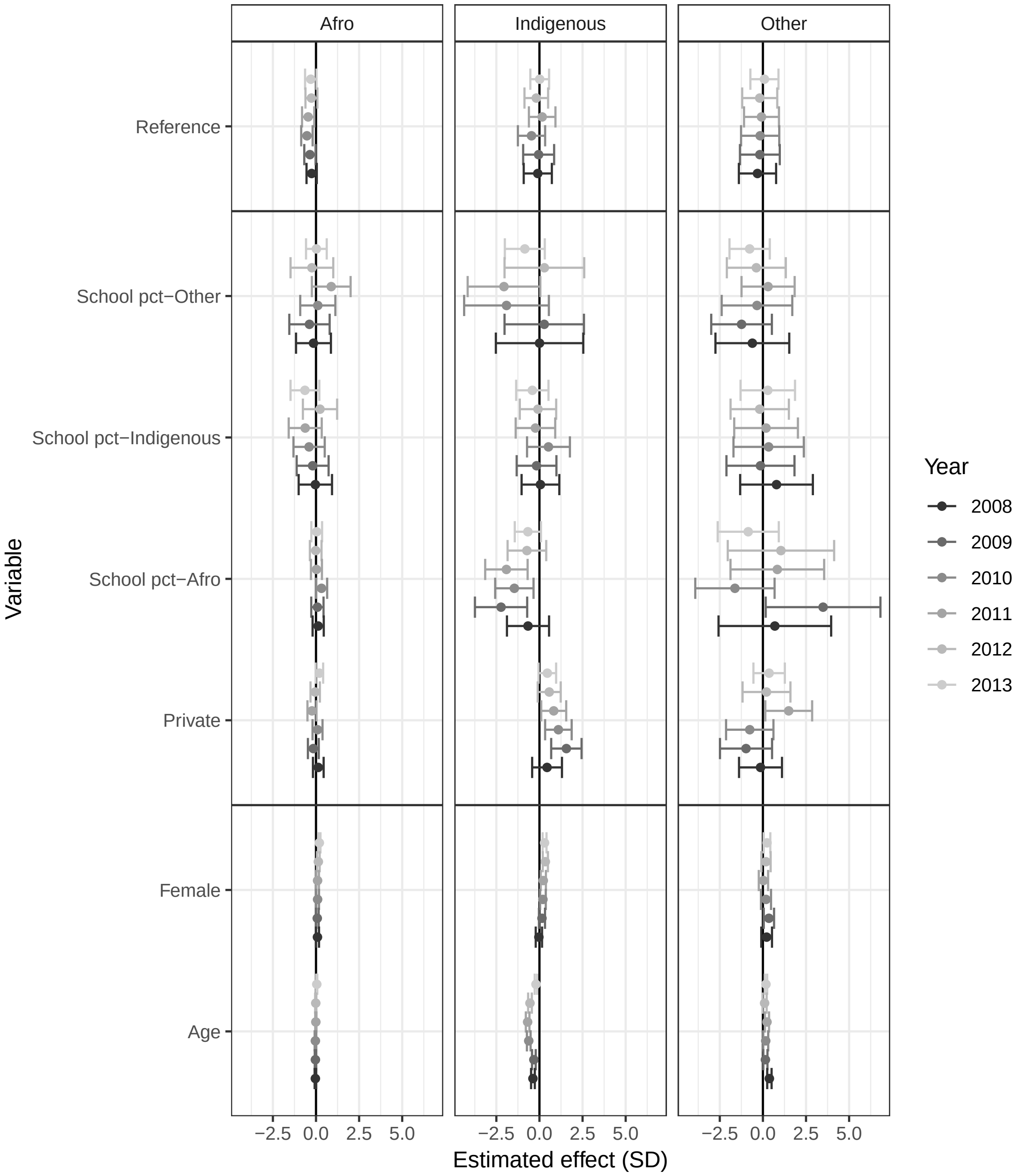

Figure A.66: Estimated interaction effects and 95% confidence intervals for the LA contextual effect of ethnicity according to the three-level random intercept model including interactions (9.4) for maths achievement by year for the sample of schools observed during all years

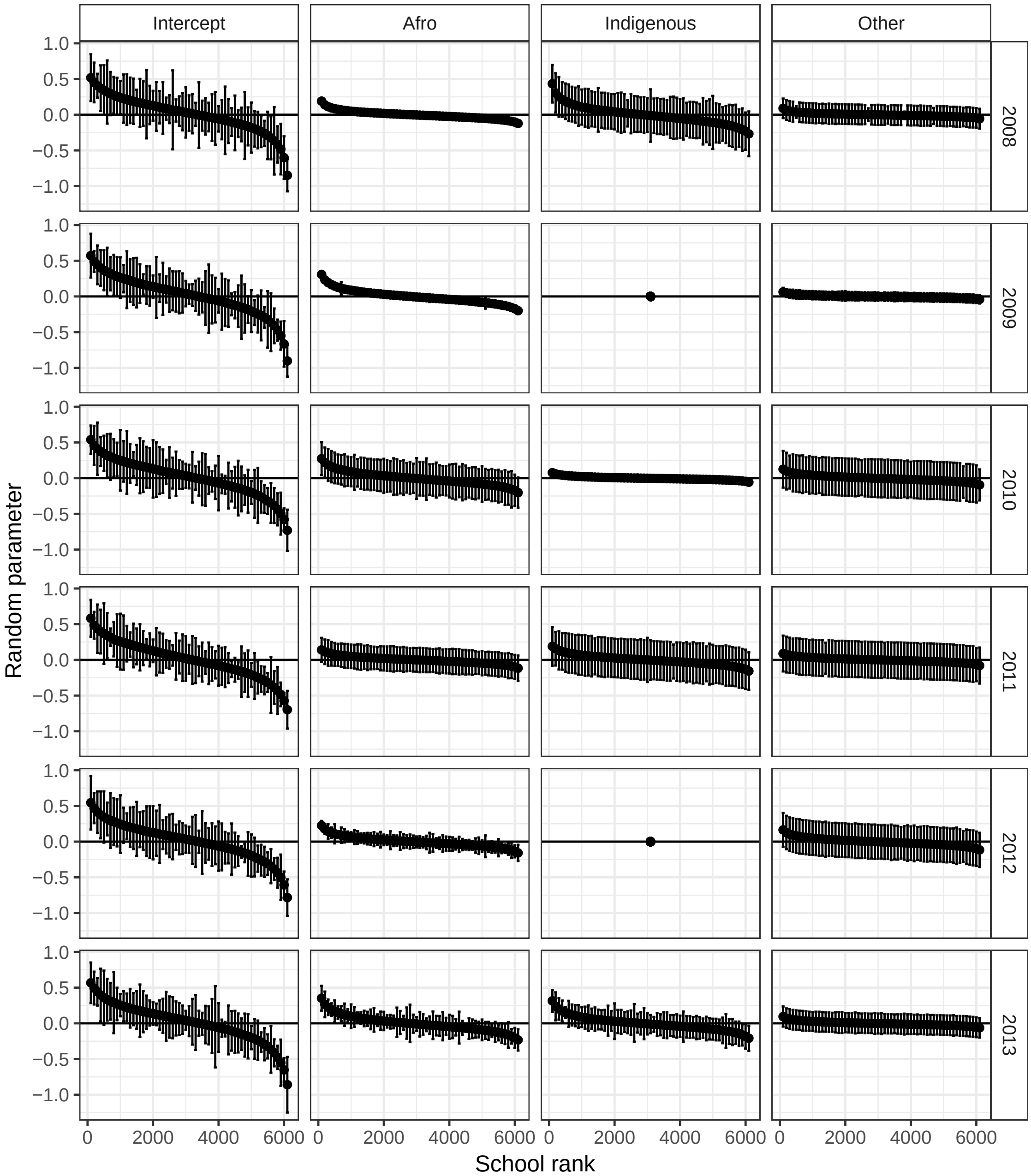

Figure A.67: Caterpillar plot for the random residual intercept and within-school ethnic achievement gap of the model given by equations (9.4) and (9.5) by year for the sample of schools observed during all years